

WISSENSCHAFTLICHE FORSCHUNGSBEITRÄGE
BIOLOGIE/BIOCHEMIE/CHEMIE

ISOLATION AND CHARACTERIZATION OF
SCR-RELATED SEQUENCES FROM THE PLATYFISH
XIPHOPHORUS MACULATUS (POECILIIDAE; TELEOSTEI)
AND AN EVOLUTIONARY ANALYSES OF THE
SRC GENE-FAMILY

Scott McGregor Robertson

Disbui. München
1989

BAND 37





Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41554827_0027

Scott McGregor Robertson

**ISOLATION AND CHARACTERIZATION OF
SCR-RELATED SEQUENCES FROM THE PLATYFISH
XIPHOPHORUS MACULATUS (POECILIIDAE; TELEOSTEI)
AND AN EVOLUTIONARY ANALYSES OF THE
SRC GENE-FAMILY**

WISSENSCHAFTLICHE FORSCHUNGSBEITRÄGE BIOLOGIE/BIOCHEMIE/CHEMIE

BAND 37

WISSENSCHAFTLICHE FORSCHUNGSBEITRÄGE
BIOLOGIE/BIOCHEMIE/CHEMIE

Herausgeberin:

Mihaela Centea-Intemann

Band: 37

CIP-Titelaufnahme der Deutschen Bibliothek

Robertson, Scott McGregor:

Isolation and characterization of src-related sequences from the platyfish xiphophorus maculatus (poeciliidae; teleostei) and an evolutionary analyses of the src gene-family / Scott McGregor Robertson.

- Prien am Chiemsee : Intemann, 1989

(Wissenschaftliche Forschungsbeiträge Biologie, Biochemie, Chemie

Bd. 37)

Zugl.: München, Univ., Diss., 1989

ISBN 3-926323-41-8

NE: GT

WG: 32

Verlag:

Andreas S. Intemann und Christoph C. Intemann (GbR), Prien am Chiemsee,
Wissenschaftlicher Verlag und technisch-wirtschaftliche Beratung,
8210 Prien, Neugartenstr. 17, ☎ (08051) 61159.

Druck: unikopie München

**ISOLATION AND CHARACTERIZATION OF SCR-RELATED SEQUENCES FROM
THE PLATYFISH XIPHOPHORUS MACULATUS (POECILIIDAE; TELEOSTEI)
AND AN EVOLUTIONARY ANALYSES OF THE SRC GENE-FAMILY**

**Inaugural-Dissertation
zur Erlangung des Doktorgrades des Fachbereichs
Biologie
der Ludwig-Maximilians-Universität
München**

**Vorgelegt von
Scott McGregor Robertson
aus Auckland, New Zealand**

München, 1989

1. Berichterstatter: Prof. Dr. H. Schmiegel

2. Berichterstatter: Prof. Dr. H. Jäckle

Tag der mündlichen Prüfung: 03.02.89

ACKNOWLEDGMENTS

I thank Manfred Scharf, who has played a major helping role, both scientific and personal, throughout this work. In addition, I extend my thanks to Prof.'s H. Schmieger and H. Jaeckle of the LMU, Munich for agreeing to serve as examiners.

I am grateful to the students, technicians and staff of the Genecenter, Martinsried for an enjoyable 3-4 years. Many contributed in some way directly or indirectly to this work and, while not being able to name all, I would like to acknowledge a few in particular: Sabine Andexinger for technical assistance with a smile; Frits for technical tips and brain-storming; Yogi for good cheer and support; and Jochen and Mike, friends who were there when they were needed - thanks guys!

Lastly, and in a well-deserved position of honor, I thank Agnes, without whom this may not have happened.

I dedicate this thesis to my parents, who instilled in all their children a love of learning and thinking. Thanks Ma!

I would like to thank the Medical Research Council (Canada) and the Max-Planck-Society for financial support during part of this work.

TABLE OF CONTENTS

	<u>PAGE</u>
I. INTRODUCTION	1
II. MATERIALS AND METHODS	19
1) Animals	19
2) Preparation of genomic DNA	20
3) Plasmids/Hybridization probes	21
4) Southern blots/Hybridizations	21
5) Slot blots	22
6) Preparation of genomic libraries; Library screening	22
7) 'Giant plaque' analyses	23
8) Subcloning into plasmid vectors	23
9) Lambda DNA preparation	24
10) Plasmid DNA preparation	25
11) Sequencing	28
12) Sequence analysis	28
13) Evolutionary analysis	29
III. RESULTS	30
1. Isolation and characterization of genes from the fish <i>src</i> /TK gene family.	30
2. Investigations of the role of the c- <i>src</i> gene in the crossing-conditioned melanomas of <i>Xiphophorus</i> .	53
3. Evolutionary aspects of the <i>src</i> /TK gene family.	61

IV.	DISCUSSION	77
	1. Fish <i>src</i> /TK gene family.	77
	2. Investigations of the role of the c- <i>src</i> gene in the crossing-conditioned melanomas of <i>Xiphophorus</i> .	81
	3. Evolutionary aspects of the <i>src</i> /TK gene family.	90
V.	SUMMARY	104
VI.	ZUSAMMENFASSUNG	106
VII.	REFERENCES	108
VIII.	Appendices	
	1.	130
	2.	131
	3.	133
	4.	134
	5.	135
IX.	LEBENS LAUF	143

ABBREVIATIONS

BC	-backcross generation
bFGF	-basic fibroblast growth factor
bp	-basepairs
c-	-cellular
cAMP	-cyclic adenosine monophosphate
<i>C. elegans</i>	- <i>Caenorhabditis elegans</i>
cpm	-counts per minute
CSF-1	-colony stimulating factor 1
dATP	-deoxyadenosine triphosphate
DNA	-deoxyribonucleic acid
EDTA	-ethylenediaminetetracetic acid, disodium salt
EGF	-epidermal growth factor
F ₁ /F ₂	-1st/2nd filial generation
FGF	-fibroblast growth factor
Fig.	-Figure
<i>H. attenuata</i>	- <i>Hydra attenuata</i>
IGF	-insulin-like growth factor
kb	-kilobase pair
LiCl	-lithium chloride
LMP	-low-melting-point
<i>mdl</i>	-macromelanophore determining locus
min(s)	-minute(s)
mRNA	-messenger RNA
Myr	-million years
M	-molar
mm/mg	-milli(10 ⁻³)meter/milligram
µm/µg/µl	-micro(10 ⁻⁶)meter/microgram/microliter
ng	-nano(10 ⁻⁹)gram
OD	-optical density
onc	-oncogene
PAM	-percent accepted mutations
PDGF	-platelet-derived growth factor



PEG	-polyethylene glycol
pfu	-plaque forming unit
pg	-pico(10 ⁻¹²)gram
PM	-plasma membrane
-R	-receptor
RFLP	-restriction fragment length polymorphism
<i>Rmdl</i>	-regulatory locus of md1
RNA	-ribonucleic acid
SDS	-sodium dodecyl (lauryl) sulfate
ser/thr	-serine/threonine
Sd/Sp/Sr	-macromelanophore spot phenotypes: spotted dorsal/spotted sides/striped sides
SSC	-NaCl/sodium citrate solution: 0.15M NaCl, 0.015M Na-citrate, pH 7.0
STET	-8%(w/v) sucrose, 0.5%(v/v) Triton X-100, 50mM Tris.Cl(8),50mM EDTA
TE	-10mM Tris.Cl (8.0), 1mM EDTA
TELT	-2.5M LiCl, 50mM Tris.Cl(7.5), 62.5mM EDTA(8.0), 0.4% Triton X-100
Tris	-tris(hydroxymethyl)amino methane
UEP	-Unit Evolutionary Period
UMGCG	-University of Wisconsin Genetics Computer Group
v-	-viral
v/v	-volume/volume
w/w	-weight/weight
w/v	-weight/volume
<i>X. maculatus</i> /	
<i>X. mac</i>	- <i>Xiphophorus maculatus</i>
<i>X. helleri</i> /	
<i>X. hell</i>	- <i>Xiphophorus helleri</i>
<i>Xhomeobox</i>	- <i>Xiphophorus</i> homeobox-containing gene
<i>Xsrc</i>	- <i>Xiphophorus</i> c-src gene
<i>Xyes</i>	- <i>Xiphophorus</i> c-yes gene

I. INTRODUCTION:

The work to be presented in this dissertation concerns a particular family of proto-oncogenes - the *src* gene family - of the poeciliid fish *Xiphophorus maculatus*. Specifically, three major areas are dealt with: firstly, the isolation and characterization of *src* -related genes from the *Xiphophorus* genome; secondly, the possible role of the *c-src* gene in melanoma formation in these fish; and thirdly, certain aspects of the evolutionary history of this gene and gene family. This introduction is therefore arranged into three sections. The first section will present a brief history of the major concepts that form the foundations of modern-day cancer genetics as well as an overview of the modern evidence for a genetic component in cancer. The second section will summarize our present knowledge concerning oncogenes and cancer as well as presenting the evidence supporting a role for recessive phenomena in tumorigenesis. The final section of the introduction will present the *Xiphophorus* melanoma system as an extremely useful animal system for the study of cancer genetics, including both dominant and recessive oncogenes. This animal model has been utilized as the experimental system throughout most of the work presented in this thesis.

A) CANCER GENETICS

In 1820, W. Norris, reporting on the occurrence of malignant melanomas in a family concluded: "The facts incline me to believe that this disease is hereditary" [219]. This is one of the earliest published reports suggesting an hereditary factor in cancer. It was more than 40 years later that Gregor Mendel performed and reported his crossing experiments and not until almost a further ten years that both DNA and chromosomes were discovered. It was around this time that Boveri was formulating what is now called the Chromosome Theory of Cancer, based primarily upon his observations of the consequences of abnormal mitoses in sea urchin eggs. This theory, first suggested by Boveri in 1902 and in more detailed form in 1914 [40], states that: 1) cancer is a cellular problem; 2) that "... typically every tumor arises from a single cell, [i.e. the now so-called clonal theory of cancer]; 3) the cancerous cell "... contains, as a result of an abnormal process, a definite and wrongly combined chromosome constitution", and (4) "This is ... the cause of the tendency to rapid cell proliferation, which is passed on to all the descendents of the primordial cell."

At around the same time, additional clinical and experimental observations were beginning to be reported consistent with a role for hereditary factors in cancer. For example, in 1905 Owen reported the occurrence of retinoblastoma in the offspring of patients cured of the same tumor [229a], and in 1910 de Gouvea reported basically the same observations [63]. But it was at the turn of the century that experimental cancer genetics really began in earnest, in most cases utilizing the mouse as the experimental system. Between 1900 and 1920 a number of groups, central among them those of Tytzer, Bashford, Murray, Little and Strong, reported on the role of hereditary factors in mouse tumors, including familial aggregation of cancer and multi-factorial tumor susceptibility [reviewed in 133,179].

Between 1920 and approximately 1945 there were a number of reports based upon a genetic analysis of cancer, including: 4-5 genes involved in tumor susceptibility, cancer presenting as a form of 'dominance', susceptibility to adenocarcinoma linked to another locus, cancer as a simple Mendelian recessive, a sex-linked gene influencing tumor susceptibility, tumors due to a simple Mendelian dominant gene, and finally reports that neither single dominant nor single recessive factors are involved [reviewed in 133,179].

The major arguments used today in support of a hereditary or genetic role in cancer are briefly presented below.

The occurrence of cancers, often of one particular type, within particular families, is highly suggestive of a genetic factor playing a role in the tumor formation [36,38]. For example, during the first decade of this century it was observed in mice that mammary cancer occurred in certain families more often than in others, and in addition that females in whose ancestry mammary cancer occurred not further back than grandmothers were more likely to develop cancer than those in whose ancestry it was more remote [133,179]. Examples in humans include: familial polyposis coli [44], familial dysplastic nevus syndrome [117], multiple endocrine adenomatosis types I, II and III [24], Lynch syndrome (cancers of colon and endometrium) [185], Li-Fraumeni syndrome [178] and retinoblastoma (see next section).

A second observation in favor of a genetic component of cancer is the association of increased risk for cancer with over 200 single-gene disorders in man. The most well-known examples of these are. Xeroderma pigmentosum, an autosomal recessive defect in repair of UV-damage to DNA associated with an extremely elevated susceptibility to skin cancer; ataxia telangiectasia, an autosomal recessive disorder associated with chromosomal abnormalities and lymphoid neoplasms; Fanconi's anemia, an autosomal recessive disorder associated with chromosomal damage and a high leukemia incidence, and Bloom's syndrome, again an autosomal recessive disorder also associated with chromosome breakage and a high cancer incidence (often acute leukemias) [reviewed in 124].

Thirdly, it has been observed for some time that many cancers are associated with chromosomal abnormalities (eg. chronic myelogenous leukemia and the Philadelphia chromosome [256,337]; Burkitt's lymphoma and translocations involving a particular cellular proto-oncogene - *c-myc* [157a])

Fourth, as most clearly shown by Ames, the high correlation between a compounds carcinogenic activity and its mutagenic activity (most often attributable to its effect upon DNA) implies that the carcinogenic process at some point involves damage to or modification of the DNA [8].

Fifth, the fact that through inbreeding with selection strains of laboratory mice can be produced that develop specific kinds of cancer of specific incidences generation after generation is held as good evidence for a genetic component underlying these tumors [133]. The emergence of these inbred strains of mice differing in their tumor incidence rates in the 1930's and 1940's led to a number of experiments alluded to earlier where a high incidence strain was crossed with a low incidence strain and F_1 , F_2 and backcross analyses performed in the hope of determining the mode of Mendelian inheritance of the tumor susceptibility or tumor resistance, or the number of genes involved [133]. This then led to a further series of experiments whereby two low incidence

strains or strains not selected for tumor incidence at all were crossed, and the tumor incidence and tumor spectrum in the F_1 , F_2 and backcross generations compared to the two parental strains [133]. These experiments illustrated that the crossing of two different low tumor incidence strains of mice could lead to an F_1 generation with a higher incidence and greater variety of tumors than in either parental strain, whereby multiple tumors were found more frequently and some tumors were observed in the F_1 animals that were not found in either parent. Possibly the best system demonstrating a crossing-conditioned carcinogenic process, and thereby the direct involvement of genetic factors in the tumor formation, is the swordtail-platyfish melanoma system [113], outlined in detail below.

Lastly, the demonstration that purified DNA was able, under certain circumstances, to directly transform cells to the malignant phenotype [169,209,278] is perhaps the most direct proof of the role of genes in this process (these experiments are discussed further below).

There can be little doubt, however, that environmental factors play a central role in most human cancers, and that truly hereditary cancers make up only a minority of the neoplasms seen in man [184,208,302]. Indeed, familial and/or occupational-related exposure to certain environmental carcinogens may be at least in part responsible for familial aggregation of some cancers - eg. familial mesothelioma and asbestos exposure; familial aggregation of leukemias following exposure to benzene. However, Knudson has suggested that all types of cancer seen in man can have a truly hereditary form which are however difficult to detect due to their low occurrence relative to environmentally induced tumors [161]. The association of smoking and lung cancer, and UV exposure and skin cancer amongst caucasians, leave little doubt concerning the overwhelming importance of environmental factors for human cancer. However, cancer is nevertheless a genetic disease at the cellular level, in that genetic damage and/or modification underlies the neoplastic transformation of the cell, be that a somatic or a congenital event, and this phenotype is relatively stably transmitted to all progeny cells. In addition, there is considerable evidence that susceptibility towards cancer contains a considerable hereditary component. That a change of some sort occurs at the level of the gene(s) during neoplastic transformation is now widely accepted, and the next section outlines our current knowledge concerning some of the genes thought to be involved in this process.

B) GENES IMPLICATED IN TUMOR INDUCTION/PROGRESSION

i) 'Dominant' viral/cellular oncogenes

In 1965 Fried isolated a conditional mutant affecting the ability of polyoma virus to transform cells in culture [87], but the 'cancer genes' of DNA tumor viruses would prove extremely difficult to define, since they are intimately involved in viral expression and replication, and because DNA tumor viruses kill the cells in which they replicate [32]. The initial identification and isolation of genes acting in a dominant fashion to transform cells (referred to as oncogenes) was originally achieved by virologists working with the RNA transforming viruses, the retroviruses. Retroviruses are not hampered by these disadvantages inherent to the DNA tumor viruses, and in 1970 Martin obtained conditional mutations in the

viral gene (*v-src*) responsible for neoplastic transformation by Rous sarcoma virus [192]. In 1970 Duesberg and Vogt also reported the occurrence of naturally occurring deletions that remove part or all of *v-src* and which are non-transforming [75], and these two observations led eventually, within the next ten years, to the cloning and sequencing of the *v-src* gene [58,65,270,306].

In the early seventies Huebner and Todaro formulated the 'virogene-hypothesis' [139,309], which stated that covert viral oncogenes (also referred to simply as *v-oncs*) were present in the genome of normal cells - i.e. that proviruses consisting of endogenous retroviruses complete with oncogene exist in normal cells. By 1976 the group of J. M. Bishop and H. Varinus had succeeded in preparing a probe specific for *v-src* sequences [296], and were therefore in a position to be able to directly test the 'virogene-hypothesis'. *v-src*-related sequences were indeed detected in DNA from normal, uninfected cells, but the initial evidence was not consistent with these sequences being part of an endogenous retrovirus [297]. A number of additional observations now invalidate the 'virogene-hypothesis': 1) cellular sequences related to the viral oncogenes (also called proto-oncogenes) are found located at a constant genetic locus in all members of a species, whereas endogenous retroviral genes are very diverse in terms of number, distribution and positioning [140]; 2) *v-oncs* and proto-oncogenes are not isogenic [76,314], with the most obvious (but by no means only) difference being the presence of introns in all proto-oncogenes (with the exception of *c-mos*); 3) no proto-oncogene has yet been found in healthy individuals within or even linked to a complete or defective provirus of an endogenous retrovirus [100], and 4) there is also no evidence that transcription from cellular proto-oncogenes is even coordinated with the expression of endogenous retroviruses. Therefore, the cellular homologues of the viral oncogenes are normal constituents of the cellular genome, presumably performing some normal function(s), but capable, under the appropriate conditions, of inducing neoplastic transformation. The identification in 1964 of the Harvey murine sarcoma virus [130], in 1966 of the Moloney murine sarcoma virus [202] and in 1970 of the Abelson murine leukemia virus [1] led to studies of their viral oncogenes (*v-Ha-ras*, *v-mos* and *v-abl*, respectively) which confirmed the basic conclusions from the *v-src* studies. In addition, these viruses demonstrated, in conjunction with experiments involving the rescue of transformation-defective mutants of *v-src*, that the oncogenes present in the acutely transforming retroviruses were the result of a recombination event (or events) between a chronic leukosis virus (containing no oncogene) and the normal cellular proto-oncogene, whereby the proto-oncogene results in being 'activated' in some way [140].

When it was realized that the acutely transforming retroviruses had acquired a normal cellular sequence that was responsible for the cellular transformation, the question immediately arose whether the normal cellular gene possessed latent transforming ability - i.e. whether these cellular genes were ever involved in tumor formation. Alternatively, the act of acquisition by the virus and maintenance thereafter in the viral genome (resulting in a very high mutation rate due to the low fidelity of the virally-encoded reverse transcriptase enzyme) had so changed the cellular sequence that a totally new, transforming function resulted. Specifically, it was wondered whether a sufficiently high level of expression of the normal cellular gene, perhaps in an inappropriate cell type or at the wrong time, would lead to cellular transformation. In the vast

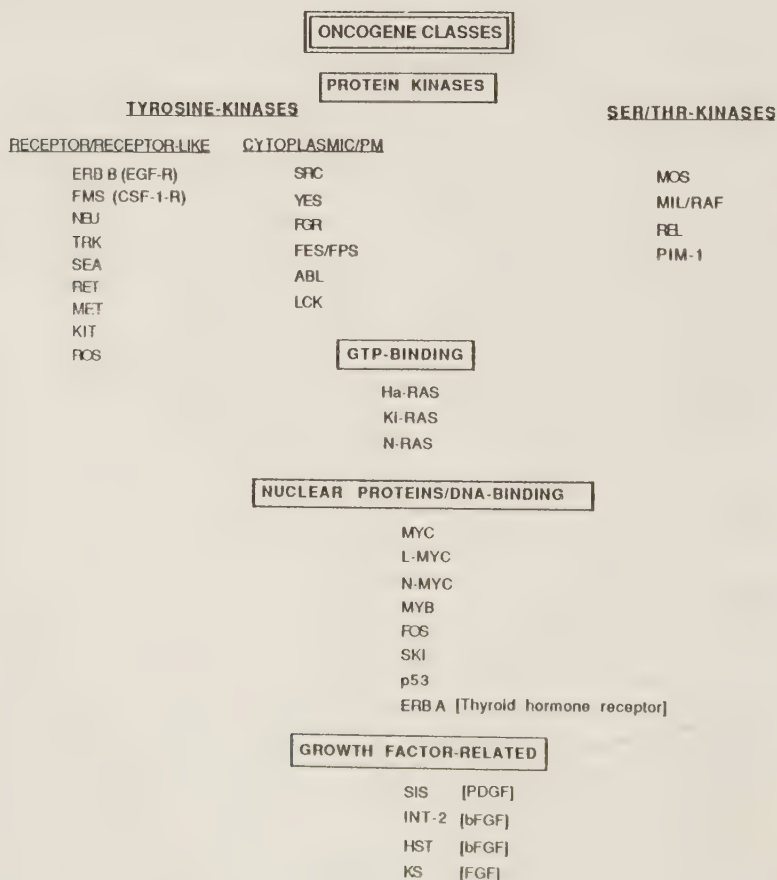
majority of cases it has been possible to show that the retroviral oncogene is a greatly altered form of the cellular gene, often involving rearrangements and/or deletions in addition to numerous point mutations. In several cases it has been possible to directly link one or more of these specific changes to the transforming function [54,142,177]. In addition, in most cases where the normal cellular proto-oncogene has been tested for a transforming function in culture, they have tested negative [85,145,235,276], although *c-mos* and *c-ras*, when expressed to sufficiently high levels, can transform immortalized mouse cells (eg. NIH 3T3) [229]. A link was soon found however between the role of the viral oncogenes in tumor induction and a possible role for the cellular proto-oncogenes in human carcinogenesis. After being able to demonstrate the transformation of cells in culture by DNA derived from human tumors or tumor cell lines it was shown that the human gene most commonly involved was the cellular homologue of the previously identified viral *ras* oncogene [66,232]. Such transforming versions of the cellular proto-oncogenes, as well as any other cellular sequences capable of neoplastically transforming cells are termed cellular oncogenes (c-oncs).

Since the original identification and isolation of the first retroviral oncogenes, a large and ever-expanding array of oncogenes, proto-oncogenes and related genes have been identified by a number of different approaches [for a review of v- and c-oncs see 33]. By far the most productive route towards the identification of transforming genes, aside from characterizing new retroviral isolates, involves DNA transfection into immortalized cell lines in culture. Total high molecular weight DNA from a particular tumor or tumor cell line is transfected into an immortalized (but non-neoplastic/tumorigenic) cell line and the presence of an activated oncogene assayed by in vitro parameters (i.e. focus formation; anchorage-independent growth) and/or in vivo parameters (i.e. tumorigenesis in the nude mouse).

Based upon function (either theoretical or proven), subcellular localization, sequence homology, etc., it has been possible to place many (although certainly not all) oncogenes into a small number of classes, and to deduce potential relationships between oncogene classes [32,312]. By far the majority of the classifiable oncogenes encode protein products with protein kinase activity [143,271]. The majority of these protein kinases are specific for tyrosine residues, as opposed to the more abundant serine- and threonine-protein kinases in the normal cell [144,272]. Most of these oncogenic proteins are located at, or just under, the plasma membrane. The second class of oncogenes encode guanine nucleotide binding proteins - the *ras* gene family (*Ha-ras*, *Ki-ras* and *N-ras*) - which are also often associated with the underside of the plasma membrane [183]. The third class of oncogenic proteins are found located predominantly in the nucleus whereby a DNA-binding and/or transcriptional activation activity have been demonstrated for several members (eg. *fos*, *myc*, *myb*, *jun* /AP 1) [79]. Lastly, there is a group of oncogenic proteins related to various growth factors (eg. *sis* - PDGF; *int-2* - FGF) [64,70,114].

The tyrosine-kinase class of oncogenes can be further divided into two separate sub-classes: those derived from normal cellular genes that encode growth factor receptors (eg. *v-erb B* /EGF receptor; *v-fms* / CSF-1 receptor), and those derived from cellular genes whose normal product is localized solely within the cytoplasm. The first subclass (i.e. receptor-related) are similar to the normal cellular proteins to the extent that they are transmembrane proteins with an intracellular tyrosine kinase domain. However, in most cases the extracellular ligand-binding

domain is severely truncated such that ligand binding can no longer occur [27,194,311]. In addition, truncations of varying extent have also occurred at the extreme carboxy-terminus of the protein next to the tyrosine kinase domain and various numbers of point mutations may also be involved [54]. An exception to this general format is the *neu* oncogene (HER2), whose ligand-binding domain is intact and the protein is apparently activated towards a transforming function by a single amino acid substitution within the transmembrane domain [20]. The second subclass of oncogenic tyrosine kinases (i.e. cytoplasmic) also usually contain short truncations of the cellular gene as well as point mutations and are predominantly found in association with the underside of the plasma membrane [168] or other membranes [253]. In several instances this subcellular localization has been shown to be necessary for transforming function [57].



These data suggest a molecular parallel to the cellular characteristics of the cancer cell: the cancer cell apparently disregards or misinterprets normal extracellular signals regulating cell division and cell differentiation. The general pathway of signal recognition and transmembrane signal transduction, intracellular signal transduction and nuclear signal interpretation/cellular response are represented by the major oncogene classes. The two subclasses of tyrosine kinases both place the kinase domain directly underneath the plasma membrane, well positioned to serve as a signal transducer from the membrane (where any extracellular signals must first be encountered) into the cytoplasm. The *ras* family of oncogenes appear to be homologous to the G-proteins and are conceivably playing a similar role in signal transduction from the membrane to the cytoplasm [39]. The oncogenic proteins displaying a nuclear localization include several genes known to play a role in the very early response of cells to various normal growth factors (eg. to EGF and PDGF) [17,170]. In addition, the fact that several oncogenic proteins mimic normal growth factors suggests another route to short-circuit this signal transduction pathway, namely by the transformed cell itself producing its own growth factor to which it can then respond (the so-called autocrine model of cell transformation) [288].

The normal human genome includes an astonishingly high number of genes encoding protein kinases (presently numbering in the dozens and potentially numbering in the hundreds), leading Hunter to suggest that protein phosphorylation is an extremely ancient and central mode of gene regulation that has been duplicated and modified many times by the cell [143]. It has indeed been known for some time that the activity of many proteins in eukaryotic cells is regulated by reversible covalent phosphorylation [257]. Based upon specifics of sequence homology and gene makeup, the two tyrosine kinase sub-classes of oncogenes (i.e. receptor-like or cytoplasmic) can be shown to be further composed of a number of gene families, which together make up the tyrosine kinase super-family [143]. For example, within the receptor-like sub-class are found the EGF-receptor gene family (EGF-R/ERB B, *neu*/ERB B2), the insulin receptor gene family (insulin-R, IGF-1-R, *c-ros*, *c-met*, *c-trk*), and the PDGF-receptor gene family (PDGF-R, CSF-1-R/*c-fms*, *c-kii*). Within the cytoplasmic sub-class are found the *abl* gene family (*c-abl*, *c-arg* [169a]), the *fms* gene family (*c-fms*, *c-flt* [195a]), and the *src* gene family [143].

The *src* gene family is presently composed of seven known members: *c-src* (isolated from human, mouse, chicken, Xenopus and fish - see below) [16,193,305,307], *c-yes* (human and fish - see below) [303], *c-fyn* (human) [156,275], *c-fgr* (human) [153,217], *c-hck* (human and mouse) [159,243,338], *c-lyn* (human) [334], and *c-lck* (mouse) [191,319]. This gene family is not yet fully defined and additional members probably exist - eg. the chicken *c-ikl* gene [301]. The *src*, *yes* and *fgr* genes have been detected originally as retroviral oncogenes (in the chicken Rous sarcoma, the chicken Y73 and the Gardner-Rasheed feline sarcoma viral isolates respectively) [155,192,211], the mouse *c-lck* gene was identified due to its rearrangement and overexpression in the T cell lymphoma cell line LSTRA [191,319], and the other members of the gene family have been isolated by way of sequence homology. In addition to a high degree of sequence similarity between all members, the *src* gene family is more precisely defined by the fact that all members (based upon the genomic sequence available for the various members) share

the same exon/intron configuration [217,338]. That is, the gene family members all have the same number of exons, the corresponding exons are all of the same size (with minor exceptions; see discussion), and the introns are all located in identical locations. This exon/intron configuration is not shared by the other gene families within the cytoplasmic sub-class of tyrosine kinases- eg. *c-abl* and *c-fps* [252,322].

The role of tyrosine kinases in general and *c-src* / *v-src* in particular in the regulation of cellular proliferation and differentiation has been extensively studied. That *c-src* and other members of the *src* gene family are playing a role, particularly regarding the regulation of cellular differentiation, is evidenced by the following observations: a number of groups have reported elevated *c-src* expression in neuronal cells with little or no replicative potential [56,90,262,285]; introduction of *c-src* expression constructs into PC 12 cells (phaeochromocytoma cell line) induces neurite outgrowth [6]; elevated *c-src* expression is associated with terminal differentiation of myeloid cells [21a]; expression of *c-lck* is entirely restricted to lymphoid cells [191], and the expression of *c-hck* is observed predominantly in granulocytes with no replication potential and a very short in vivo half life (<12 hours) [243,338]. A number of studies have identified potential substrates for the *c-src* and the *v-src* proteins but these studies have not yet brought about a significant advance in our understanding of the normal or transforming function of the *src* protein [reviewed in 323].

Sufficient information is slowly becoming available that it is possible to make some tentative conclusions regarding the evolution of certain cellular proto-oncogenes. First, it is apparent that not all proto-oncogenes are found in all eukaryotes. Although there have been occasional unsubstantiated claims, the consensus of opinion at the moment is that the proto-oncogenes identified to date are not found in the plant kingdom. The *ras* family of proto-oncogenes is extremely ubiquitous, with *ras* genes having been isolated from such diverse organisms as *Dictyostelium* (1 gene) [248], yeast (2 genes, plus one related gene) [62,91], *Drosophila* (3 genes) [215] and humans (minimum of 4 genes- *Ha-ras*, *Ki-ras*, *N-ras* and the related *R-ras*) [182,280]. In fact, a *ras*-related gene encoding a GTP-binding protein has even been identified in bacteria [5]. These genes are extremely highly conserved over evolutionary time in terms of both sequence, gene organization, and function - so much so that a human *ras* gene can complement a defective yeast gene [154]. This suggests very strongly that the *ras* genes are performing a function central to the normal cellular physiology of perhaps every animal cell. By comparison to the near ubiquitous distribution of the *ras* genes, it would appear that a number of other proto-oncogenes are found in a more restricted subset of organisms. For example, *src* appears to be restricted to multicellular metazoans [262] (consistent with a proposed role in intercellular communication and neuronal differentiation - see below), and *fos* appears to be restricted to the vertebrate lineage [210,258,260] (consistent with a proposed role in bone development and, as activated oncogene, in the development of osteosarcomas). Second, as already suggested by the *ras* gene family data above, the protooncogene gene families show an increase in the number of family members over evolutionary time. Third, the parallel in increasing complexity of both protooncogenes and cell/tissue complexity during evolution is consistent with the observation of few, if any, neoplasms in evolutionarily very primitive organisms and a general increase in both

the rate of occurrence and the variety of tumors in more advanced organisms [60,150,214].

Since the initial observation that the kinase domain of the viral *src* protein was related to the catalytic chain of mammalian cAMP-dependent protein kinase [21], it is apparent that most (and possibly all) eukaryotic protein-serine/threonine and protein-tyrosine kinases show sequence similarities in the kinase domain [144,312]. This has led to the suggestion that these genes have all arisen from a single primordial gene [143]. Somewhat surprisingly, very little is known concerning the evolution of the protein kinase gene super-family and in particular the *src* gene family, although in this case its derivation from an ancestral gene by way of gene duplications and divergence is generally accepted. Tyrosine kinase activity and/or *src*-related sequences have been detected in a wide variety of organisms, including all vertebrates looked at, sea urchin (Echinodermata) [244], *Drosophila* (Arthropoda) [279], cuttle fish (Mollusca) [M. Scharlt; pers. comm.], *Caenorhabditis elegans* (Nematoda) [98], *Hydra attenuata* (Coelenterata) [266] and sponge (Porifera) [22,262].

ii) 'Recessive' cellular oncogenes (Antioncogenes)

A number of observations, however, led to reservations concerning the oncogene field as it has been outlined above [reviewed in 158,162,292]. First, it was clear very early on, mainly from clinical studies, that carcinogenesis was a multi-step process [46,81,82], and this was confirmed and reproduced in tissue culture and whole animal model systems of chemical carcinogenesis [81,203]. The fact that one activated oncogene appeared responsible for the entire neoplastic phenotype, as evidenced by the viral oncogenes and the DNA transfer experiments, was inconsistent with this notion. More recent experiments with oncogene transformation have helped to clarify this apparent discrepancy. The first observation was that primary cells in culture were not transformed to the full neoplastic phenotype by a single activated oncogene, but instead required at least two complementing genes [172]. It would appear that one of the complementing oncogenes immortalizes the primary cells, while the second oncogene completes the neoplastic transformation. The NIH 3T3 cells, which are most often used in the cell culture assay for activated oncogenes, are already immortalized and are readily transformed by one activated oncogene of this second complementing class. Second, the introduction of activated oncogenes into the germline of transgenic mice often resulted in reproducible tumor formation, which was also initially interpreted as supporting a dominant neoplastic transforming function by a single oncogene [122,188,299]. However, closer examination of these systems showed consistently that although the presence of these oncogenes is necessary to produce the noted effects it is generally not sufficient - that is, in virtually every case there were clear indications that secondary events were also necessary for oncogenesis in these transgenic animals [123].

There were also a number of observations involving apparently activated oncogenes in neoplastically transformed cells that were not easily interpreted in terms of a single dominant oncogene underlying the neoplastic event. The simple co-cultivation of transformed cells with normal cells can in certain cases lead to the reversible normalization (i.e. loss of the transformed phenotype) of the malignant cells [200,300]. This same phenomenon has been observed in experiments involving the introduction of malignant cells into embryos. The introduction of mouse

teratocarcinoma cells into the blastocoel of an early embryo resulted in the 'normalization' of the malignant cells and led to the contribution of these cells to all major lineages of the developing embryo, including the germ line in some individuals [201]. Because the teratocarcinoma cells are believed to represent the neoplastic equivalents of totipotent early embryonic cells, these results were interpreted as resulting from the return of these cells to the appropriate embryonic environment where they came under, and responded correctly to, the normal regulatory environmental cues. This is consistent with the experimental method of generating teratocarcinomas, which is to graft early-stage embryos to non-uterine sites - i.e. to totally disrupt the normal embryonic environment for these cells [298]. These initial observations regarding teratocarcinoma cells in chimaeric embryos have been significantly extended by Pierce and coworkers. They have been able to repeat these sorts of experiments for neuroblastoma cell lines and for melanoma cell lines, and in both cases could show that only a particular embryonic environment or 'field', defined precisely in terms of both time and position within the embryo (and coinciding with the normal development of the neuroblasts or the melanoblasts), would result in the normalization of the introduced malignant cells [95,242,326]. Sachs has been able to show that a similar but distinct embryonic field also exists for myeloid leukemia cells introduced into early mouse embryos [104]. These results demonstrating the reversibility of the transformed phenotype both in culture and in the embryo are again inconsistent with transformation by dominant acting oncogenes. Lastly, in the course of investigating the developmental potential of both embryonic and adult frog nuclei by injecting them into enucleated fertilized frog eggs, several groups tested nuclei from malignant cell lines (eg. renal carcinoma), and observed the development of tumor-free post-hatching tadpoles [69].

The fact that only certain cell types are transformed by an activated oncogene, whether this occurs in tissue culture, in a retrovirus-infected animal or in transgenic animals, although the oncogene itself may be expressed at high levels in a number of other cell types [34a,114a], raises the possibility that the susceptible cells lack the appropriate suppression system whereas in the resistant cells this system is intact. That is, these observations raise the possibility that neoplastic transformation can result from the absence or functional inactivation of a particular gene or genes - i.e. through a genetically recessive process. The first experimental evidence consistent with a recessive mutation in somatic cells giving rise to a tumor came from somatic cell hybrid studies from Harris's group, first reported in 1969 [128]. In a series of experiments (since confirmed in a number of laboratories), Harris and coworkers were able to show that the malignant phenotype was usually suppressed in hybrid cells formed between malignant cells and normal cells [129,259,291]. It was further shown by Harris's group and others (predominantly the group of Stanbridge) that the initial non-malignant hybrid cell could 'revert' to malignancy and that this phenomenon was correlated with the loss from the hybrid cell of a particular chromosome or chromosomes which derived from the normal parent cell [80,293]. Stanbridge has shown by this sort of chromosome segregation analysis that the maintenance of human chromosome 11 in somatic cell hybrids is sufficient to suppress the malignant phenotype of HeLa cells and Wilm's tumor cell lines [290]. More recently these workers could show that introduction of a normal chromosome 11 on its own was capable of suppressing the malignancy of the HeLa and the Wilm's cells [325]. Further experiments from

Stanbridge's group involving suppression of the transformed phenotype in certain malignant x malignant cell hybrids suggest that not just one but several suppression systems exist in the cell, which are able to genetically complement one another in the hybrid cell [294].

In addition, it has been demonstrated that transformed cells containing known, supposedly dominant-acting activated cellular and viral oncogenes, can also be suppressed when fused with certain normal cells. Examples include fusions of nontransformed rat or mouse cells and RSV-transformed rat kidney cells [78]; 3T3 and SV40-transformed BALB/c mouse lines [136,137]; and flat revertants of Ki-MuSV-transformed NIH-3T3 cells and other 3T3 cell lines transformed by either *ras* or *les*-containing viruses [218]. This suppression was observed to occur at different levels - in the first case suppression was at the level of transcription (neither *v-src* mRNA nor pp60^{v-src} protein were detected), whereas in the other two examples suppression was post-translational (T antigen or p21^{ras} were present at normally transforming levels in the suppressed cells).

The best evidence for neoplastic transformation resulting from recessive phenomena comes from the studies, both clinical, cytological and molecular, of retinoblastoma, a childhood ocular tumor. In 1971 Ohno suggested in a general model of carcinogenesis that malignancy was much more likely to result from recessive genetic changes than dominant ones and that these recessive mutant alleles might be detected by the loss of the unaffected gene on the homologous chromosome [223]. Also in 1971 Knudson presented the same sort of model for a specific tumor type - retinoblastoma [160]. Recent studies have confirmed Knudson's model, and have placed the affected locus (*Rb*) at band q14 on the long arm of chromosome 13 [48,286]. The use of a well-characterized panel of deletions involving 13q14 has recently led to the isolation of a cDNA clone with properties consistent with it deriving from the *Rb* gene [89]. However, introduction of the wild-type *Rb* gene into retinoblastoma cells and their concomitant normalization has not yet been reported. There is a rapidly expanding list of other tumors which, like retinoblastoma, may be the result of recessive genetic changes and are often associated with specific chromosomal deletions. These include Wilm's tumor (and associated tumor syndromes) [167,226] and neuroblastoma [163], both of which, like retinoblastoma, are commonly childhood tumors. In addition, these childhood tumors are often part of a pathological syndrome that involves other distinct tumor types (eg. retinoblastoma patients have a greatly increased frequency of osteosarcoma in later life) [125,166,273], consistent with the proposal above from Stanbridge that several distinct suppression systems exist, each appropriate for particular cell types. In the case of Wilm's tumor, the chromosome implicated as normally containing the locus deleted in the tumor cells (chromosome 11) is the same chromosome that when introduced into Wilm's cells in culture suppresses the neoplastic phenotype of these cells. These results argue quite strongly that a 'tumor-suppressor' gene specific for kidney cells of a particular developmental stage maps to human chromosome 11.

There have emerged several models of carcinogenesis based upon these sorts of observations and which do not rely upon a strict mutational mechanism - i.e. models designed precisely to explain this reversibility of the neoplastic phenotype [134,251]. A variety of observations support the suggestion that DNA methylation may be playing a role. First, a large number of studies support the involvement of DNA methylation in the regulation of gene expression -

specifically, transcriptionally active genes have often been shown to be hypomethylated [3,49,50,72]. Second, the methylation pattern of a particular gene has been shown to be preserved over many cell generations, which could account for the level of stability observed for the transformed phenotype [126,327]. Third, a number of cellular variants that had been ascribed to classical mutational mechanisms have since been shown to be fully reversible by compounds that inhibit DNA methylation [127]. Therefore an epigenetic change at the level of DNA methylation can accurately mimic a true mutational event (i.e. one involving an actual change in the DNA sequence) and yet remain fully reversible. Lastly, there have been a small number of reports concerning hypomethylation of specific genes, including oncogenes, in a number of tumor samples [reviewed in 4].

The modern day concept of neoplastic transformation is therefore moving to a large extent away from the relatively naive idea of dominant oncogenes playing a singular role. Instead, the concept that both positive (i.e. growth stimulatory) and negative (i.e. growth inhibitory) factors are playing a role is becoming more generally accepted [114,288,289,321]. One of the earliest general carcinogenesis models to be presented based upon recessive phenomena was that of Osgood in 1957 who proposed that the absence of an inhibitor of growth normally secreted by the terminally differentiated cells could result in the transformed phenotype [227]. Very recent experiments investigating the role of the growth inhibitory factor transforming growth factor- β in both retinoblastoma and breast cancer support Osgood's proposal [195]. However, it was in fact Ribbørt - in 1895 - who was one of the first to postulate that a malignant change can originate from loss of growth restraint by adjacent tissue, enabling cells to continue proliferating and give rise to a tumor [249].

Although the majority of the studies outlined above have been performed either in human or higher vertebrate systems, these systems suffer from a number of drawbacks when one attempts a molecular genetic analysis of tumor formation. First, in the vast majority of cases the underlying genetics of the tumor formation are extremely complex - potentially a number of genetic loci are involved; environmental influences are difficult to control; long tumor latency periods are involved, etc. Second, the generation times are relatively long. Transgenic mice carrying activated oncogenes at least have the advantage of reproducible recapitulation of events leading to the tumor formation in those mice that inherit the transgene, but the system is still artificial to some extent and subject to the variability of the secondary events alluded to above that are also involved in the tumorigenic process. Studies in *Drosophila*, although in a system with extremely powerful genetics, also suffer from several drawbacks - 1) the animal is an invertebrate and there is some question regarding the analogy of *Drosophila* tumors to vertebrates and mammals, 2) insect embryogenesis and development is so different to that of vertebrates that attempts to correlate the normal function of an oncogene in *Drosophila* to man, for example, may not always succeed (eg. *int -1* and *wingless* [24b]), 3) carcinogenesis with chemicals is not possible in adult *Drosophila*, and 4) all *Drosophila* tumors occur as embryonic and larval lethals, and do not occur in the adult fly [94]. The platyfish/swordtail melanoma system [see for reviews: 9,113,151,230,268,269,315] has the advantages 1) that it is a vertebrate tumor system with clear cellular and biochemical parallels to the mammalian situation, 2) that the genetics

underlying the tumor formation, well worked out over the last fifty or so years, appear relatively straight-forward, and 3) that the tumor itself is readily detected and monitored by eye.

C) *XIPHOPHORUS* MELANOMA SYSTEM

Fish breeders have long known that crossing individuals from different species with different color patterns can lead to hybrid fish larger and much more colorful than either parent. Early in this century several fish breeders performing such crosses with various *Xiphophorus* species noticed "cancer-like" growths in some of the hybrids [107,131,164]. These first reports of this phenomenon also called attention to the fact that these pigmented tumor-like growths only occurred when at least one of the original parent species was carrying a specific pigment cell spot pattern. Between 1927 and the early 1960's predominantly the group of Myron Gordon in New York determined both the genetic and the cytological basis for the tumor formation in the hybrid fish [18,107-113,165].

Platyfish and swordtails are small fishes indigenous to the Atlantic fresh water river systems of Mexico and Central America [151]. These viviparous (i.e. live-bearing) teleosts are members of the genus *Xiphophorus*, family Poeciliidae. *Xiphophorus* has been classified into a large number of species [255] which although phenotypically distinguishable and reproductively isolated can be hybridized in the laboratory, and all hybrids (with few exceptions) are fertile. All species looked at so far have 48 (2n) small chromosomes [88,221], and the genome size has been estimated at approximately 20% of the mammalian genome ($\sim 1.20\text{-}1.40$ pg; $\sim 4.5 \times 10^8$ bp) [221,267]. Sex chromosomes have not been demonstrated cytologically in these fish because the chromosomes are all extremely small and show very little morphological variation following Giemsa banding [53,281], but instead have been identified and followed genetically by studying the inheritance of a number of dominant sex chromosome-linked pigment patterns [151] - see below.

The fish display species and population specific pigmentation patterns due to cells in the dermis containing the pigments pteridine (pterinophores), guanine (guanophores) or melanin (melanocytes and melanophores) [151]. The term melanophore denotes a specialized melanin-bearing cell which is usually found in lower vertebrates, and which is capable of dispersing its melanosomes from the center of the cell to its periphery and vice versa, under hormonal and/or neural control. There are two types of melanophores found in the skin of *Xiphophorus*, the micro-melanophores which are found evenly distributed over the body of the fish and which do not exceed 100 μ m in diameter, and the macromelanophores, which are considerably larger than the micromelanophore (up to 500 μ m in diameter) [105,151]. Both types of cells can be arranged in groups to form the species and strain specific spot patterns (which may include spots composed of macromelanophores of up to 1-2mm in diameter).

In the natural populations of the platyfish (*Xiphophorus maculatus*) black-spotted color patterns are found in about 20% of the individuals (1879 of 9000 adult wild platyfish from eight different river systems) [113]. Normally, *Xiphophorus helleri* (the swordtail) never present macromelanophores in their skin, although, as with the platyfish, micromelanophores are present. Humm and Young studied the embryonic origin of pigment cells in xiphophorine fish by tissue-grafting experiments and concluded that, analogous to the situation for the mammalian

melanocyte, both micro- and macromelanophores in the fish derive from the same stem cell (melanoblast), which in turn derives from the neural crest region of the embryo during stages 10-13 (neural tube stage; embryo stages according to Tavalga [308]) [141]. Therefore the melanophore cell-lineage is: neural crest - melanoblast - melanocyte - melanophore.

When a platyfish (*X. maculatus*) homozygous for the spotted-dorsal gene (*Sd*), expressed phenotypically by the presence of one or two patches (spots) of macromelanophores in the dorsal fin, is crossed with a swordtail (*X. helleri*) without a macromelanophore pigment pattern, the F1 offspring exhibit an enhancement of the macromelanophore spot pattern, often resulting in a completely black dorsal fin [109] - see also Figure 1a. If the F1 offspring are backcrossed to the *X. helleri* parent, approximately 50% of the offspring exhibit no macromelanophores (i.e. appear swordtail-like in coloration), approximately 25% exhibit a 'mild' melanosis similar to the F1 offspring, and approximately 25% exhibit a 'severe' melanosis characterized by the development of nodular tumors on the dorsal fin and adjacent areas, often with foci of macromelanophores appearing on the caudal and pectoral fins and on the lips. Typical results of such a cross are presented in Figure 1a. The fish with the mild form of melanosis live a normal lifespan and are reproductively fit whereas the fish with the severe melanosis are clearly unhealthy and shortened in life-expectancy and are reproductively very unfit. The mild and severe forms of the melanosis have been likened by several groups to a benign and malignant melanoma respectively, and cytological, biochemical and electron microscopic studies of the fish tumors support this proposal [115,250,284]. It was first shown by Gordon that this mild versus severe division of melanoma tumor phenotype is correlated with differences in levels of differentiation of the macromelanophores, with the cells of the mild (i.e. benign) melanomas appearing more completely differentiated than those of the severe (i.e. malignant) class [112]. That the tumor-bearing BC generation fish could be divided into two groups based upon tumor severity was first proposed by Anders and co-workers [10]. This has since been independently confirmed by and large by Siciliano and co-workers [282]. This observation was important because the 1:2:1 segregation in the backcross generation suggested that two independently segregating loci were regulating the presence and severity of the melanoma (see below). That the gene or genes responsible for the pigment spot pattern were centrally involved in the tumor formation was shown in addition by the fact that the origin of the melanoma on the fishes body was strictly determined by the spot pattern phenotype used in the cross - i.e. fish with spots in the dorsal fin give rise to hybrid fish with tumors emanating from the dorsal fin, fish with spots on the body side likewise give rise to hybrid fish with tumors emanating from the body side, and so on [108].

That this apparant neoplastic transformation in the xiphophorine fish was under negative control was first proposed by Breider [41] and somewhat later also by Gordon [113]. Our current working model concerning melanoma formation in platyfish-swordtail hybrids, derived in large part from these earlier models [11], can be briefly outlined as follows: there exists a gene (or genes) in the *Xiphophorus* genome, usually located on a sex chromosome, which determines the macromelanophore cell. This gene will be referred to as the macromelanophore determining locus (*mdl*), although it has received various names from different groups (eg. color gene, macromelanophore gene, Mel [for macromelanophore determinant], and Tu [for tumor gene]). It is further hypothesized that within the fish genome are a number of genes (ranging from only a

few to many, depending upon the exact genotype of the fish) which modify the expression and/or the effect of the *mdl* gene. These will be referred to generally as $\mathbf{R}mdl$ (for regulatory) genes and can be either linked to the *mdl* locus on the sex chromosome, or can be autosomal. It is further hypothesized that due to a co-evolution of the *mdl* / $\mathbf{R}mdl$ system, a particular $\mathbf{R}mdl$ locus is strictly specific for its associated *mdl* locus. In the case of the example cross outlined above, two of these regulatory loci need to be invoked. The first is thought to be tightly linked to *mdl* [152] and to have a cell-lineage effect, determining which cell lineages capable of giving rise to pigment cells differentiate towards the macromelanophore, and in which very specifically defined parts of the body this process occurs - i.e. in this case the dorsal fin of the fish. This locus will be termed $\mathbf{R}^{Sd}$, and rather than being a structural locus is more likely to represent non-coding cis regulatory sequences for the *mdl* locus. $\mathbf{R}^{Sd}$ is considered to be an allele of one of the $\mathbf{R}mdl$ loci, in much the same way that fish with spots on their body side (the *Sp* phenotype) are thought to harbour the $\mathbf{R}^{Sp}$ allele of that same $\mathbf{R}mdl$ locus. The second $\mathbf{R}mdl$ locus is unlinked to *mdl* (i.e. is autosomal) and is thought to have its effect at the level of differentiation of the final stages of the cell lineage leading to and including the macromelanophore. For this reason the gene has been referred to as *diff* (for differentiation) [318] or *mel sev* (for melanoma severity) [204] and will be referred to here as $\mathbf{R}diff$. In the presence of $\mathbf{R}diff$ the cells differentiate beyond the stage of the actively proliferating melanocyte to the terminally differentiated and non-dividing macromelanophore. For the purposes of the model one must envisage that $\mathbf{R}diff$ has a dose-dependent effect, with a double copy of $\mathbf{R}diff$ generating more of a differentiation signal than a single copy.

Based upon these concepts of *mdl* regulation, it can be postulated that in the predominantly unspotted populations of platyfish, and presumably also in the swordtail populations, *mdl* is present and is very tightly regulated by a number of both linked and non-linked regulating loci and therefore is not expressed. However, in the spotted platyfish, mutations have occurred at the $\mathbf{R}mdl$ locus linked to *mdl* that allow macromelanophore expression in precisely defined areas of the fish body. Therefore the *mdl* phenotypes *Sp* (spots on the side of the body), *Sd* (spot or spots in the dorsal fin), *Sr* (stripes on the side of the body), etc., are thought to correspond to particular $\mathbf{R}mdl$ alleles which have arisen previously through mutation at the wild-type $\mathbf{R}mdl$ locus that is linked to *mdl* - these would be $\mathbf{R}^{Sp}$, $\mathbf{R}^{Sd}$, and $\mathbf{R}^{Sr}$ respectively. Because these animals are also homozygous for $\mathbf{R}diff$, the spots are all composed of terminally differentiated and non-proliferative macromelanophores.

When a platyfish with the spotted-dorsal phenotype (i.e. $\mathbf{R}^{Sd} / \mathbf{R}^{Sd}$; $\mathbf{R}diff / \mathbf{R}diff$ - only taking into account these two regulatory loci of the *mdl* locus, remembering that $\mathbf{R}^{Sd}$ is tightly linked to the *mdl* locus) is crossed with a swordtail (the F_1 fish become heterozygous for both loci ($\mathbf{R}^{Sd} / +$; $\mathbf{R}diff / +$). Because of the single copy of $\mathbf{R}diff$ the cells of the spots have a greater proliferative capacity leading to its greater size and superficial spreading. When the F_1 fish are backcrossed again to the swordtail parent, four different genotypes will result (again only taking into account these two regulatory loci): $(+/+; +/+)$ and $(+/+; \mathbf{R}diff / +)$ which will account for approximately 50% of the offspring and, since not containing an activated (i.e. mutated) $\mathbf{R}mdl$ /*mdl* locus, will show no spots or tumor; ($\mathbf{R}^{Sd} / +; \mathbf{R}diff / +$) which will account for

approximately 25% of the offspring and will be identical in phenotype to the F_1 (i.e. mild or benign form of melanoma due to the differentiation push of the single copy of R^{diff}); and ($R^{Sd} / +; +/+$) which will also account for 25% of the offspring and due to the total absence of R^{diff} the melanoma will be composed more of actively proliferating melanocytes than macromelanophores and will cover a larger portion of the animals body as well as being more invasive. This model is presented schematically in Figure 1b. F_2 analyses and further backcrosses to the swordtail parent give results consistent with this model. As one in effect replaces platyfish chromosomes containing genes capable of regulating the platyfish *mdl* locus by swordtail chromosomes not capable of this regulation (although the swordtail presumably contains a regulation or suppression system specific for the swordtail *mdl* locus), one major prediction of this model would be that as hybrid fish bearing benign or malignant melanoma are backcrossed to the platyfish parent, the re-introduction of the appropriate platyfish chromosomes should reconstitute the complete suppression system for the *mdl* locus. This is indeed found to be the case [11]. A further prediction from this model would be that as the swordtail also contains an operative R^{mdl} / mdl system it should be possible to observe a deregulated *mdl* locus in certain swordtail populations - i.e. a macromelanophore-derived spot pattern. This is also found in one case (eg. the Dabbed [*Db*] pattern in a particular *X. helleri* population [M. Scharlt; pers. comm.]). In the model presented above R^{diff} is analogous to the *Rb* locus in retinoblastoma. In fact, the analogy gains some strength since Phillips and Gallie have also proposed that a distinct pre-malignant state also exists in retinoblastoma that corresponds to the presence of only one active *Rb* gene [90a].

Previous studies had demonstrated a strong correlation between the degree of deregulation of the *mdl* locus (i.e. the severity of the melanoma) and the expression of the *Xiphophorus c-src* gene, as measured by immunoprecipitable tyrosine-specific protein kinase activity detected using a rabbit antiserum raised against the product of the viral *src* gene. For example, in the F_1 fish with benign melanoma there is an elevation in the *src* tyrosine kinase activity compared to the similar basal level found in both parental fish, and this kinase activity is further elevated in the backcross fish with malignant melanoma (up to approximately 6-fold [263]). In the case of the melanomas that develop following these crosses this is true whether the kinase assay is performed with tumor tissue or with brain tissue from the fish (although this elevated tyrosine kinase activity is not detected in other tissues). However, in the case of *Xiphophorus* strains where tumor formation can be somatically induced (eg. following carcinogen treatment), an elevated kinase activity is detected only in the tumor tissue [264a]. These results are consistent with the fish *c-src* gene playing some as yet undetermined role in both the somatically induced and the crossing-conditioned melanomas, and would also suggest some form of co-regulation of *mdl* and *c-src*.

D) THESIS PRESENTATION

The isolation and characterization of the *Xiphophorus* *c-src* gene would allow at least three separate lines of investigation to be followed: First, cloning of the *c-src* gene would allow the analysis of the normal patterns of expression of the gene during development and in the adult animal, which may help in analyzing the gene's normal function. Second, it would enable one to more directly approach the role of the *c-src* gene in the melanomas in the fish, both at the level of expression, regulation and *mdl/c-src* relationships. Third, the isolation of *c-src* and other *src*-related sequences from the genome of the fish would allow the first analysis of members of the *src* gene family in a lower vertebrate, which should lead to insights concerning the evolution of this particular gene family. The work presented in this dissertation describes the isolation and initial characterization of several *src*-related sequences from the *Xiphophorus* genome and how these sequences have been used to approach the questions outlined immediately above. The presentation of the results and discussion are divided into three sections, the first covering the isolation and initial physical characterization of the *src*-related clones from a *Xiphophorus* genomic library, the second presenting analyses of the role of the fish *c-src* gene in the crossing-conditioned tumor formation, and the third section presenting evolutionary implications derived from the analysis of these fish sequences and their comparison to *src* gene family members from other organisms.

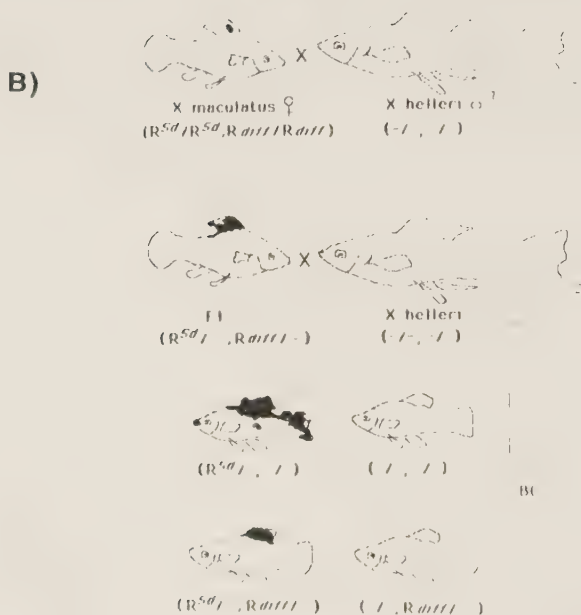


FIGURE 1

Crossing-conditioned melanoma formation in *Xiphophorus*.

A) Typical crossing scheme leading to melanoma formation in platyfish/swordtail hybrids.

A - platyfish (*X. maculatus* ; spotted dorsal (Sd) phenotype); B - swordtail (*X. helleri* ; macromelanophore free); C - F₁ hybrid ('benign melanoma'); D - BC₁-hybrids with benign melanoma; E - BC₁-hybrids with malignant melanoma; F) & G) - tumor and macromelanophore-free BC₁ segregants.

B) Genetic model for tumor formation in the platyfish/swordtail hybrids (see text for explanation)

II. MATERIALS AND METHODS

1) ANIMALS

A) *XIPHOPHORUS*

All fish employed in these studies were derived from pure wild-type populations that have been randomly inbred in the laboratory for at least 30 generations.

The following stocks were used:

Xiphophorus maculatus (Rio Usumacinta, Mexico)- macromelanophore spot pattern free [X.mac (RU)].

Xiphophorus maculatus (Rio Jamapa, Mexico)- macromelanophore spot pattern Sd ('Spotted dorsal') [X.mac (RJ)]

Xiphophorus variatus (Rio Panuco, Mexico)- macromelanophore spot patterns Li ('Lineatus') and Pu ('punctatus') [X.mac (RP)]

Xiphophorus helleri III (Rio Lancetilla, Mexico)- macromelanophore spot pattern free [X.hell;swordtail].

The following mutant and/or hybrid strains were used:

Sd^T- translocation in a *X. mac* (RJ)/*X. hell* hybrid fish of part of the *X. mac* X-chromosome containing the Sd locus to an *X. hell* autosome.

Sd^{Del}- deletion of part of the *X. mac* X-chromosome including the Sd locus.

Sd/Hell III- *X. mac* (RJ) X-chromosome (isolated) in the *X. hell* III genetic background by continual back-crossing to *X. hell* III; 50% of these fish develop benign and malignant melanoma.

Sd^T/Hell III- *X. mac* (RJ) mutant Sd^T X-chromosome in the *X. hell* III genetic background (see above); these fish develop malignant melanoma.

Sd^{Del}/Hell III- *X. mac* (RJ) mutant Sd^{Del} X-chromosome in the *X. hell* III genetic background (see above); these fish develop no melanoma, benign or malignant (no macromelanophores).

DrLi/Sr⁻- *X. mac* with two deregulated *mdl* loci (*mdl* locus Lineatus (Li) deregulated due to a chromosomal translocation; *mdl* locus Stripped Sides (Sr) deregulated due to an X-ray induced mutation).

DrLi/Ni^θ - similar to DrLi/Sr⁻ above; *mdl* locus Nigra (Ni) mutated to the deregulated Nigra extended (Ni^θ).

B) *HYDRA*

(Phylum Coelenterata (Cnidaria); Class Hydrozoa; Order Hydroida)

Hydra attenuata (*H. vulgaris*) and *Hydra viridissima* were kindly provided by the laboratory of Prof. C. N. David, Institute of Zoology, L-M-Univ., Munich, F.R.G.

C) *SPONGE*

(Phylum Porifera; Class Demospongiae; Order Monaxonida)

The fresh-water sponge *Spongilla lacustris* were collected from ponds in the vicinity of Giessen

(FRG) during spring and immediately deep-frozen (-70°C)

D) MOUSE (Balb/C); CHICKEN (White leghorn, 'Spafas' strain); *XENOPUS LAEVIS* - from commercial suppliers.

E) SHARK; LAMPREY; *Amphioxus* ; TUNICATE (*Ciona intestinalis*); CRAB - collected from the Marinebiological Research Station, Helgoland, FRG [M. Scharlt; S. Otilie and G. Hannig].

2) PREPARATION OF GENOMIC DNA [Blin and Stafford, 1976; J. Collins, pers. comm.]

DNA was prepared by a modification of the method of Blin and Stafford [36a]. Tissues that had been frozen in liquid nitrogen immediately following dissection were ground to a powder under liquid nitrogen using a pre-cooled mortar and pestle. The still frozen powder was added slowly to lysis buffer [0.1M NaCl; 0.2M EDTA, pH 8.0; 100 µg/ml proteinase K; 0.5% SDS] at 80°C, and incubated for 3 hours at 80°C with intermittent gentle mixing. Particularly when preparing DNA for cosmid libraries, extreme care had to be exercised in order to diminish the shearing forces exerted upon the DNA. At the end of the incubation, one half volume of freshly equilibrated ultra-pure phenol was added, and extraction carried out for ten minutes at room temperature. After 10 minutes, one-half (of the original) volume of chloroform/iso-amylalcohol (24:1) was added, and the extraction continued for a further 10 minutes. Following centrifugation (5-8000 x g; 10'), the upper aqueous layer was carefully removed (either using a plastic pipette with its tip snapped off and flamed if necessary to round the edges, or, alternatively, for the cosmid libraries, by very carefully pouring the solution off, avoiding contamination with the interface). This aqueous layer was gently extracted two times with an equal volume of chloroform/iso-amylalcohol, and the DNA spooled out of the final aqueous layer onto a sterile glass rod following the addition of 0.6 volumes of isopropanol at room temperature. This DNA was washed by dipping the glass rod into 70% ethanol, allowed to air dry, and carefully resuspended in TE, which can take several days. DNA concentration, molecular weight, and restriction enzyme digestability are monitored by running aliquots on a 0.6% agarose gel.

The large amounts of melanin present in the melanotic tumors interfered with the above-outlined protocol. Therefore, the following further modification was implemented. Following the three hour incubation in lysis solution, the DNA was directly precipitated by the addition of two volumes of 100% ethanol. This precipitated DNA (which contains a large amount of associated melanin) is resuspended in TE. Following full resuspension, the melanin (apparently present in the form of melanosomes) can be removed by centrifugation. The now clear supernatant above the black pellet was removed and continued through the organic extractions and was finally precipitated with ethanol.

Human peripheral blood leukocyte (PBL) DNA was a gift from W. Nellen (MPI, Martinsried).

Drosophila melanogaster (Canton S) DNA was a gift from F. Michels (MPI, Martinsried).

Tapeworm (*Diphylobothrium dendriticum*) DNA was a gift from B-G Wikgren (Åbo, Finland)

3) PLASMIDS/HYBRIDIZATION PROBES

v-src 400: Pst I fragment G; 401 bp; -76-326 (numbered from A of initiation codon of RSV *v-src* gene)

v-src 350: Pst I fragment H; 345 bp; 327-671

v-src 250: Pst I fragment I; 255 bp; 672-926 (v-src fragments from RSV SRA-2)

v-src 600: Pst I fragment F; 612 bp; 927-1538 [65]

v-yes : 1.1 kb Pst I fragment (basepairs 1650-2780 of Y73-11A) [157].

v-fgr : 850 bp Sma I/Bam HI fragment (basepairs 916-1757 of the Gardner-Rasheed feline sarcoma virus gene encoding gag-actin-fgr polyprotein) [211].

c-lck : 1.75 kb Eco RI insert of NT18, containing a mouse *c-lck* cDNA insert in pUC13 [191]

c-hck : 770 bp Pst I/Rsa I fragment of HK24, containing a 1,950 bp full-length human *c-hck* cDNA insert in pUC13 [338]

v-fps : 1.3 kb Pvu II/Sma I fragment of PRCII cloned into Sma site of pUC8 [138]

v-fms : ~1.5 kb Kpn I/Bgl II fragment from SM-FeSV cloned into Pst I site of pBR322 [73]

v-fes : ~0.7 kb Pst I fragment from 3' end of ST-FeSV [277]

v-ros : 0.75 kb Eco RI/Pvu II fragment from UR2 in Eco RI/Pvu II site of pBR322 [212]

v-abl : ~2.2 kb Bgl II/Hind III fragment from A-MuLV cloned into pBR322 [101]

v-erb B: [333]

v-mos : [247]

All fragments to be used for radio-labelling were purified from low-melting-point agarose gels and finally purified over a NACS 52 prepac column (BRL). Nick-translations [250a] were performed using a kit from Amersham and according to suppliers instructions. Alternatively, random oligonucleotide primed labelling [83a] was employed. Specific activities in the range of 3×10^8 - 2×10^9 cpm (Cherenkov)/ μ g were routinely achieved.

4) SOUTHERN BLOTS/HYBRIDIZATIONS

Following size fractionation in an agarose gel, DNA fragments were transferred, usually overnight, to nylon membranes (GeneScreen or GeneScreenPlus - NEN; according to suppliers recommendations) by alkaline transfer [247a], essentially as described. For genomic Southern blots a 15 minute treatment of the gel with 0.25 N HCl preceded the actual blot. For lambda and plasmid clones this step was usually omitted.

Hybridizations can be classified into three general categories based upon the stringency involved in the actual hybridization and/or the washings of the membranes prior to exposure to x-ray film. The hybridizations were all performed at 42°C, with the stringency being varied by altering the relative amount of formamide in the hybridization mix: 35% formamide = reduced stringency; 40% = standard stringency (i.e. moderate stringency conditions normally employed throughout this work for non-homologous hybridizations); 50% formamide = high stringency (employed for hybridizations with homologous probes). In addition to formamide, the prehybridization/hybridization mix contained: 1M NaCl, 5X Denhardtts (1X Denhardtts = 0.02%

each of polyvinylpyrrolidone, Ficoll and bovine serum albumin), 50mM Tris.Cl (pH 7.5), 1% SDS, and usually (but not always) 100 µg/ml sonicated and denatured calf thymus DNA. The normal wash parameters are presented below:

	<u>SSC (X)</u>	<u>SDS (%)</u>	<u>TEMP (°C)</u>
reduced stringency	2.0	1.0	40-50
standard stringency	0.5-1.0	1.0	50-60
high stringency	0.1-0.2	1.0	65-70

Washed filters were exposed to Kodak X-omat AR film with one intensifying screen (DuPont Cronex III) at -70°C for periods of time ranging from several minutes (certain clone hybridizations) to 1-2 weeks (certain genomic Southernns).

5) SLOT BLOTS [Beltz, G. A. *et al*, 1983]

Slot blots were prepared essentially as described [24a] using a Schleicher and Schuell slot-blot manifold II. Following washing in water for 1 hour, the filter/membrane was washed 3-4 times with 1M ammonium acetate before application of the DNA. Care was taken to quantitate the various plasmid DNA samples by multiple side-by-side analyses in agarose gels and then corrections made for the differing vector and insert sizes to ensure an equimolar loading of the v-onc sequences. The DNA was denatured in 0.3 N NaOH for 10 minutes and placed on ice; when needed, it was diluted with an equal volume of cold 2M ammonium acetate. This was then applied to one of the wells of the slot-blot manifold which was maintained under gentle suction. Immediately prior to application of the sample a small volume of 1M ammonium acetate was applied to the well, and following the application of the sample the well was washed with two more aliquots of 1M ammonium acetate. When all samples have been applied to the membrane, the membrane was washed again with 1M ammonium acetate under suction, air-dried, and baked if needed. Hybridization of the slot-blot filters was carried out as described above (see 4).

6) PREPARATION OF GENOMIC LIBRARIES; LIBRARY SCREENING.

Genomic libraries were prepared in the lambda vector EMBL 4 [89a] by essentially standard procedures. High molecular weight genomic DNA was partially digested with Sau 3A under conditions that give rise to the highest relative representation of digestion fragments in the 18-22 kb range. This region was cut out from a preparative low melting point agarose gel, the gel slice melted, and the DNA purified by passage over a NACS 52 prepac column (BRL). The purified and concentrated insert DNA was ligated to Bam HI digested lambda EMBL4 DNA (the 'stuffer' fragment having been removed by separation in a LMP-agarose gel and the vector arms being purified as outlined above) under conditions that favor the formation of intermolecular hybrids [188a]. In vitro packaging was performed with self prepared packaging extracts from bacterial strains BHB2688 and BHB2690 as described [188a]. The packaged libraries were

plated on the P2-lysogenic strain WL95 - EMBL vectors allow direct selection of recombinants on the basis of the $\text{spl}^+/\text{spl}^-$ phenotype (i.e. recombinants can form plaques on a P2 lysogenic host while wild-type EMBL phage cannot).

Ten 150mm petri-plates each carrying approximately 40-50,000 plaques (i.e. approximately 4.5×10^5 plaques in total, which corresponds to about 10-12 haploid genome equivalents) were used for screening the genomic libraries. Duplicate filters from each plate were prehybridized and hybridized at 42°C in batch in a large glass petri plate shaking at 60-70 revolutions per minute (to inhibit the filters from sticking one to another). As much prehybridization solution as possible was used, and for hybridization a minimum volume was used that nonetheless guaranteed that all filters were maintained wet without sticking to one another. Hybridization conditions were standard (see 4) above). Positive plaques were picked and put through two more cycles of plating, hybridization and picking to ensure their positive nature and their clonality. Plate lysate stocks of the lambda clones were prepared by standard methods [188a].

7) 'GIANT PLAQUE' ANALYSES [Beltz, G. A. *et al.* 1983]

An ordered array of large, uniform plaques was prepared as described [24a]. Briefly, bacterial lawns were prepared by mixing 0.25 ml of 10mM MgCl_2 -10mM CaCl_2 , 0.5 ml of a fresh bacterial overnight culture (usually WL 87), and 6.5 ml of top-agarose. This mixture was poured onto 150mm agar plates and left at room temperature for ten minutes to harden. A very small aliquot ($\sim 0.5 \mu\text{l}$) of each phage stock was then spotted onto the hardened top-agarose in a non-symmetrical ordered array. The small spot of liquid was left at room temperature for 30 minutes to ensure absorption into the agar, the plates then inverted and incubated overnight at 37°C. Uniform large plaques of $\sim 6\text{mm}$ develop. The plates were chilled for at least 1 hour before filter-lifts are made.

8) SUB-CLONING INTO PLASMID VECTORS

a) Sub-cloning in low melting point (LMP) agarose [Crouse, G.F. *et al.* 1983]

The protocol of Crouse *et al.* [57a] was followed without modification, except that for very small fragments (eg. <250 bp) a 3% Nu-Sieve (FMC) gel was utilized. The exact ligation conditions varied according to size and concentration of fragments involved, nature of the restriction enzyme sites involved, and whether one of the DNAs had been dephosphorylated with calf intestinal phosphatase (CIP). A typical ligation mixture would include: 20-40 ng of linearized, phosphatased vector (normally $\sim 3\text{ kb}$); 0.5-5ng insert DNA (normally in the range of 300-700 bp), 5 units of T4 DNA ligase, 1X ligation buffer (see Maniatis *et al.* 1982), and a final volume of ~ 250 -400 μl . Ligation was normally carried out at 12°C overnight. During this time the agarose normally hardens, and was therefore vortexed to liquify it before transformation of competent bacteria (see below).

b) Preparation of competent bacteria [Hanahan, D., 1986]

Competent bacteria were prepared according to Hanahan [123a] protocol 3, and frozen in aliquots at -70°C. Following freezing and thawing these bacteria normally gave between 2×10^6 and 5×10^7 colonies per μg supercoiled plasmid DNA (pUC 18).

c) Size-fractionation of fragment of interest prior to ligation,

The aim of this modification was to fractionate a complex mixture of DNA restriction fragments, and then to assay separate fractions for the presence of a particular sequence which can be detected with a hybridization probe. The source DNA (lambda clone, plasmid subclone, or purified fragment) was digested with an enzyme or pair of enzymes that generates a large number of small fragments. The DNA digest was separated in a 0.6% LMP-agarose gel, and the gel sliced into fractions that differ by approximately 200-300 basepairs. Each gel slice was melted and an aliquot run on a second gel (mini-gel, 20 minute run-time) which was then blotted (1-3 hours) or, alternatively, dried. Following hybridization with the appropriate probe, that gel slice which gave the strongest hybridization signal is remelted and ligated as described above (see 7a).

d) Generation of deletion series following subcloning.

Following subcloning into the appropriate plasmid vector it was on occasion necessary to generate a deletion series to aid in sequence analysis. This was performed with Exo III and Exo VII as described by Yanisch-Perron *et al* [334a]. Exo III is an exonuclease specific for 5'-overhang or blunt restriction enzyme termini and which digests only one of the two DNA strands. When a circular DNA molecule is digested with two different restriction enzymes, one of which generates a 5' or blunt terminus and the other a 3'-overhang, then subsequent incubation with Exo III for varying periods of time results in progressive deletion of one of the DNA strands from the 5' or blunt terminus. Incubation with Exo VII, an exonuclease specific for single-stranded DNA, removes this 'tail' left following the Exo III digestion, and the resultant shortened DNA molecule can be recircularized and transformed into competent bacteria. A series of plasmid mini-preparations and digestions allows the ordering of a series of clones with different degrees of deletion.

9) LAMBDA DNA PREPARATION

a) Mini-prep. [Helms, C. *et al*. 1985]

Lambda mini-preps were prepared essentially as described by Helms *et al* [131a]. The basis for this protocol is that phage particles from a plate or liquid lysate are concentrated and purified by binding to a Whatman DEAE-Cellulose (DE 52) column prior to purification of lambda DNA by standard methods [188a].

b) Maxi-prep.

An overnight culture of host bacteria (WL 87 or LE 392) was prepared in NZ medium [188a]. The OD_{600} [$1 \text{ OD}_{600} \sim 8 \times 10^8$ bacteria/ml] was measured and 4×10^9 bacteria transferred to a separate 15ml Falcon tube per maxi prep. Cells were harvested by centrifugation

at 4000g at room temperature for 5-10 minutes. After removal of the supernatant the bacterial pellet was resuspended in 1.5 ml SM buffer (see Maniatis et al, 1982). Approximately 6×10^7 phage particles (pfu's - plaque forming units) from a recently titred phage stock were added, and incubation carried out at 37°C for 20 minutes. This mixture was added to 100mls of pre-warmed NZ medium in a 1 liter Erlenmeyer flask and incubated at 37°C with hard shaking (i.e. 220-250 rpm in a bacterial shaker) until lysis was evident (~6-9 hours). 5mls chloroform were added, and shaking continued at 37°C for a further 20-30 minutes. The lysate was centrifuged at 7000 rpm at 4°C for 10 minutes in the GS-3 rotor(Sorval) and the supernatant carefully removed to a new 250 ml centrifuge bottle (leaving the chloroform behind). RNase A and DNase I were added to final concentrations of 10µg/ml each, and the mixture incubated at 37°C for 30-60 minutes. An equal volume (i.e. ~ 100 mls) of 20% PEG (w/v), 2M NaCl in SM buffer was added and left at 0°C for at least 1 hour (or overnight). Following centrifugation at 8000 rpm at 4°C for 20 minutes in the GS-3 rotor, the supernatant was carefully removed and discarded. The pellet (containing the phage particles) was resuspended in 12 mls SM buffer overnight and then centrifuged at 8000 rpm at 4°C for 2-5 minutes to remove insoluble material. To the transferred supernatant in a fresh 50 ml Falcon tube were added: 60 µl 20% SDS, 125 µl 0.5M EDTA, and incubation carried out at 68°C for 15 minutes. The mixture was extracted carefully with one volume buffer-equilibrated phenol, then with one volume of phenol/chloroform/isoamylalcohol (25:24:1), and finally with one volume of chloroform/isoamylalcohol (24:1). One volume of iso-propanol was added to the final aqueous phase, which was then left at -70°C for 20 minutes. Following centrifugation at 10,000 g at 4°C for 10 minutes (SS34 rotor), the resultant pellet was washed with 70% ethanol, and without letting the pellet dry completely (in which case the DNA may become partially resistant to certain restriction enzymes), resuspended in 200-300 µl TE (10mM Tris.Cl, pH 8.0; 1mM EDTA). DNA yield and digestibility of the DNA were assayed in an agarose gel following resuspension and restriction enzyme digestion.

10) PLASMID DNA PREPARATION

a) MINI-PREPS

i) Boiling method [Maniatis, T. et al, 1982]

One ml of an overnight culture was spun in an Eppendorf table-top centrifuge at top speed for 15 seconds. The supernatant was carefully removed (using a Pasteur pipette that has been pulled to a fine tip in a flame) and the bacterial pellets were loosened by vortexing two tubes together (so that the two tubes 'clatter'). The loosened pellets were carefully resuspended in 100 µl TELT buffer (or alternatively STET buffer) [188a]. 10 µl of freshly prepared lysozyme solution (10 mg/ml in water) was added and mixed. The tubes were then immediately placed in a 100°C waterbath for one minute and then placed on ice for 5 mins. Following centrifugation for 8 mins at room temperature (Eppendorf table-top centrifuge; top speed), the pellet of bacterial debris and chromosomal DNA was removed with a sterile (autoclaved) toothpick. The DNA (and RNA) in the supernatant was precipitated with two volumes of ice-cold 100% ethanol. Following centrifugation and washing with 70% ethanol, the nucleic acid pellet was dried for one minute

under vacuum. The dried pellet was resuspended in 50 μ l TE (pH 8.0). 5 μ l was usually sufficient for an analytical restriction enzyme digestion and gel.

ii) Alkaline Lysis [Birnboim, H. C., 1983]

0.5 ml of an overnight bacterial culture was spun in an Eppendorf table-top centrifuge at top speed for 15 seconds. The supernatant was carefully removed and the bacterial pellets were loosened as outlined above. Each pellet was resuspended in 0.1 ml of lysozyme solution by vortexing immediately after each addition. The tubes were then placed on ice for 5 mins. Following the addition of 0.2 ml of alkaline SDS (at room temperature) the tubes were mixed gently by inversion (2-3X) and returned to ice for 5 mins. 0.15 ml of high salt solution was added (at room temperature), mixed as above, and incubated on ice for 15 mins. Following centrifugation at room temperature (Eppendorf; top speed) for two minutes, 0.35 ml of the supernatant was transferred to a new 1.5 ml tube. 0.9 ml ice-cold ethanol was added, mixed, and the tube placed at -20°C for 15 mins. Centrifugation at top speed for one minute pelleted the nucleic acids, which were then dissolved in 0.1 ml acetate-MOPS. 0.2 ml ethanol was added to precipitate nucleic acids, then centrifugation and precipitation repeated. Following washing with 70% ethanol, the final pellet was resuspended in 40 μ l water or CDTA-Tris (or TE). Approximately 1/5 of each sample was sufficient for an analytical restriction enzyme digestion.

b) MAXI-PREPS

i) Boiling/CsCl

One liter of an overnight culture was spun at 4,000 rpm for 5 minutes in the GS-3 rotor. The supernatant was poured off, the tubes inverted, and the last of the supernatant removed with a Kleenex. The pellet was resuspended carefully in 16 mls TELT buffer (or STET buffer) [188a] using a circular rotating platform, and then transferred to a 250 ml Erlenmeyer flask. Two mls of lysozyme solution (10 mg/ml in water; prepared fresh) was added, the flask heated in the flame of a Bunsen burner for 5-10 seconds, and then placed immediately in a boiling water-bath and shaken for 1 minute. The flask was then placed on ice for 5-8 minutes. The lysate (which is extremely viscous) was transferred to a 50.2 Ti ultra-centrifuge tube (using a plastic funnel to aid in transfer), and centrifuged at 30,000 rpm at 15°C for 30 mins. The supernatant was transferred to a 50 ml Falcon tube and extracted twice with phenol/chloroform/iso-amylalcohol (25:24:1) by vortexing for 1 minute. The phases were separated by centrifugation at 5,000 rpm for 10 mins at room temperature in a Hettich or Christ Minifuge. The aqueous phase was transferred to a new 50 ml Falcon tube and 0.8 volumes of iso-propanol added. The tube was placed on ice for 15 mins, then centrifuged at 4-5,000 rpm at 4°C for 10 mins (Hettich or Christ Minifuge). The pellet was washed with 70% ethanol and allowed to dry slightly, then resuspended in 3 mls TE by vortexing. The solution was transferred to a 15 ml Falcon tube containing 4.5 g CsCl and 0.5 ml ethidium bromide (10 mg/ml in water), and vortexed until everything was resuspended. Following centrifugation at 6000 rpm for 10 mins (Christ Minifuge), the supernatant was transferred to a VTi65 quick-seal tube which was then filled with water and/or 50% CsCl (w/w in water) so that the final weight of the tube plus contents was 9.35-9.8 g. Centrifugation was carried out at 55,000 rpm at 20°C for 12-14 hours. The (lower) plasmid

DNA band was removed either by dripping out from the bottom of the tube, or by withdrawing the band with a syringe. The solution was transferred to a fresh VTI65 quick-seal tube which was filled again as above and centrifuged at 55,000 rpm at 20°C for 5 hours. The plasmid band was removed as above. Repeated extraction (i.e. at least two more times following apparant clearing of the aqueous phase) with 50% CsCl-saturated isopropanol was performed to remove the ethidium bromide. Two volumes of water, and 0.6 volumes (final) of 100% iso-propanol were added, and the mixture placed on ice for 15mins or longer. DNA was pelleted by centrifugation at 10,000 rpm at 4°C for 10 mins (SS34 rotor). The pellet was washed with 70% ethanol, then with 96% ethanol, and allowed to dry only slightly before resuspensionnnnn in 300-500 µl TE. DNA concentration was determined by running aliquots of DNA on an agarose gel with known standards.

b) ALKALINE LYSIS/GLASS POWDER [Marks, M. A. et al, 1982]

One liter of an amplified overnight culture was centrifuged at 6000g for 10 mins at 0°C and washed with 50 mls H₂O. The bacterial pellet was carefully resuspended in one ml of lysozyme buffer and then 9 mls of lysozyme buffer containing 10 mg of lysozyme was added and mixed. Following incubation of the tubes on ice for 30 mins, 20 mls of alkaline SDS (at room temperature) was added and stirred gently with a sterile glass rod until nearly homogeneous. The lysate at this point is extremely viscous. Following incubation of the tubes at 0°C for 10 mins 15 mls of high salt solution was added and the lysate stirred a little more vigorously than before until a coarse white precipitate forms (and the viscosity disappears). The tubes were incubated at 0°C for 30 mins and then centrifuged at 12,000g for 10 mins at 0°C. The supernatant was then transferred to a new tube and total nucleic acids precipitated by the addition of two volumes of cold 100% ethanol. The tubes were placed at -20°C for 20' and then centrifuged as previously.

The pellet was dissolved in 5 mls of acetate-MOPS and total nucleic acids again precipitated with two volumes of cold 100% ethanol. Following centrifugation the pellet was dissolved in two mls of water, the precise volume was determined, and an equal volume of LiCl solution added. During an incubation at 0°C for 15 mins a precipitate forms which was removed by centrifugation at 12,000g for 10 min at 0°C. The supernatant was removed to a new tube and heated at 60°C for 10mins, during which time a small precipitate often formed. This was removed by centrifugation as above. Two volumes of cold 100% ethanol were added to the supernatant and the tubes placed at -20°C for 15 mins. Centrifugation was carried out as above and the pellet resuspended in 2.5 ml acetate-MOPS. The ethanol precipitation was then repeated. The final pellet was resuspended in 2 mls glass powder loading buffer [190a] and combined with 20 mls of a stock suspension of glass powder (prepared by grinding glass-fibre filters by hand with a mortar and pestle) in loading buffer and mixed gently at room temperature for 15 mins. The slurry was poured onto a 0.45 µm Millipore or Nalgene filter and gentle suction applied. The DNA was washed with approximately 120 mls of loading buffer, taking care not to allow the powder to dry (which leads to irreversible binding of the DNA to the glass powder). The DNA was eluted with 24 mls of elution buffer [190a] under constant gentle suction (i.e. 1-2 mls/minute) and all interstitial fluid was removed by applying full suction at the end of the elution. The now-purified plasmid DNA was precipitated by the addition of two volumes of cold 100% ethanol and incubation at -20°C for a minimum of one

hour (or over-night). Following centrifugation the pellet was resuspended in 4 mls of 0.1M NaOAc; 50mM Tris.Cl (pH 8.0) and precipitated with two volumes of cold ethanol. This final pellet was dissolved in 1 ml of CDTA-Tris (or TE) and stored at 4°C over a drop of chloroform. Concentration was determined as above (see i).

11) SEQUENCING [Sanger, F. et al. 1977]

All sequencing was done by the chain termination method of Sanger et al. [260a] using dideoxynucleotides as chain terminators and [³⁵S]-dATP as radioactive label. Published modifications to this protocol that were incorporated into the sequencing protocol followed throughout this work include: the use of alkali-denatured double-stranded closed circular plasmid templates, prepared either as mini- or maxi-preps [53a,240a]; the use of [³⁵S]dATP in place of [³²P]dATP as the radioactive label, with concomitant sharper bands on the autoradiograms [27a,226a]; the use of an aluminium plate to help disperse heat evenly during the gel run, and thereby reduce 'smiling' artefacts of the bands detected in the autoradiogram [327a]; running the sequencing reaction with Klenow polymerase at elevated temperature to minimize artefacts due to secondary structure of the DNA template [103a]; and the covalent bonding of the sequencing gel to one of the glass support plates, reducing the handling of thin and fragile gels [93a].

The sequencing mixes, gel mixes etc. were prepared from standard published protocols [eg. 200a], and during part of this work deoxy- and dideoxynucleotide stock solutions were obtained from a commercial supplier (Pharmacia).

The more recent sequencing experiments have been performed exclusively with the Sequenase kit supplied by United States Biochemical (USB). This utilizes a modified T7 DNA polymerase enzyme which exhibits high processivity, low 3'-5' exonuclease activity, and high speed. This has the practical advantages that the sequencing reactions are extremely rapid, even when carried out at room temperature, and the enzyme is less sensitive to the secondary structure of the template DNA. All mixes and stock solutions are supplied with the kit and were used according to the suppliers instructions.

12) SEQUENCE ANALYSIS

All DNA sequence analysis was performed with the University of Wisconsin Genetics Computer Group (UWGCG) program package [67a] on a Digital Equipment Corporation MicroVax computer at the MPI for Biochemistry (Martinsried, FRG) Computer Center. Specifically, the programs GAP, BESTFIT and COMPARE/DOTPLOT were used in the direct comparison of two or several sequences, in all cases at the program default values for all parameters. WORDSEARCH was used to compare a particular DNA or amino acid sequence against available sequence data banks.

BESTFIT finds the best segment(s) of similarity between two sequences by using the 'local homology' algorithm of Smith and Waterman [283a]. GAP aligns sequences along their entire length (as opposed to finding the best segments of similarity as in BESTFIT) using the algorithm of Needleman and Wunsch [212a]. GAP adds gaps to either sequence to achieve the maximal number of matches. COMPARE creates a file of the points in common between 2 sequences which can be

plotted out with DOTPLOT. COMPARE compares 2 sequences for similarity in every possible frame by looking through a window (the size of which the operator can set - 'window') for a certain number of matches (which also can be set by the operator - 'stringency').

13) EVOLUTIONARY ANALYSIS

All distance matrices used in the construction of phylogenetic trees were corrected according to Dayhoff, 1978 to take into account, on a purely statistical basis, those changes that have taken place in the past but that are no longer observable, based upon a mutation data matrix derived from the available sequence data and compiled by Dayhoff [61].

Phylogenetic trees were constructed with the PHYLIP package of programs (Version 3.0, April, 1987) provided by Joseph Felsenstein. The programs utilized were FITCH and, predominantly, KITCH, which are both designed for calculating possible evolutionary relationships based upon a distance matrix of discrete characters. FITCH produces a tree with branch lengths unconstrained; KITCH on the other hand assumes that all tip species are contemporaneous and that an evolutionary molecular clock is operating, which means that branches of the tree cannot be of arbitrary length, but are constrained so that the total length from the root of the tree to any species is the same.

III. RESULTS

1. ISOLATION AND CHARACTERIZATION OF GENES FROM THE FISH *SRC* /TK GENE FAMILY

Family of *src* -related sequences in fish genome

When a *Xiphophorus* genomic Southern blot is probed under moderate stringency conditions (see Materials and Methods) with a viral *src* fragment derived from the highly conserved kinase domain (Pst I 600 bp fragment F [65] of Rous sarcoma virus, see Appendix Fig. 1), a relatively complex pattern of hybridizing bands of varying intensities is observed (Fig. 2a, lanes 1 and 3). When the same filter is re-probed under identical stringency conditions with a viral *yes* fragment also from the kinase domain, a very similar pattern of hybridizing bands is detected (Fig. 2a, lanes 2 and 4). The bands in this case are also of varying intensities, and individual bands hybridize most strongly with either the *v-src* or the *v-yes* probe. The comparison of these relative intensities suggests that the fish genome contains sequences more closely related to *v-src* and other sequences more closely related to *v-yes*. These hybridizations have been repeated with probes from the kinase domains of the *v-fgr*, human HCK and mouse *c-lck* genes with results consistent with the fish genome containing separate sequences closely related to each probe (Figure 2b and c). These results indicate that, as assayed by DNA hybridization, the fish genome contains a family of *src* -related sequences, analogous to the situation in humans [143].

Isolation of *src*-related clones

In order to isolate these *src* -related sequences from the fish genome, a genomic library of *Xiphophorus maculatus* DNA was prepared and screened with the *v-src* kinase domain probe under identical conditions as used for the genomic Southern.

High molecular weight DNA was prepared from testes of *Xiphophorus maculatus* (Rio Usumacinta), which contains a strongly regulated *mdl* locus (i.e. no macromelanophore spot pattern) as well as from a strain containing two deregulated *mdl* loci (DrLi/Ni⁰ - see Materials and Methods). The DNA was prepared by a modification of the procedure of Blin and Stafford, partially digested with Sau 3A, and the 15-20 kb fraction was purified from a 0.6% LMP agarose gel by passing the melted gel slice over a NACS 52 column (BRL). Purified insert was ligated to Bam HI digested EMBL 4 arms ('stuffer' removed by gel purification) and packaged using self-prepared extracts (see Materials and Methods). Recombinants were selected on the basis of the spi+/spi- phenotype (i.e. recombinants can form plaques on a P₂ lysogenic host while wild-type phage cannot). The original library sizes (i.e. prior to amplification on the P₂ lysogenic strain) were approximately 0.9×10^6 plaque forming units (pfu) for the Rio Usumacinta and 1.2×10^6 pfu for the DrLi/Ni⁰ libraries. Given that the *Xiphophorus* genome is approximately 4.5×10^8 basepairs [267] these libraries represent approximately 30-35

genome equivalents and have a greater than 99.9% probability of total genome representation. All further results presented here concerning the *Xiphophorus src* -related clones involve only the *X. maculatus* (Rio Usumacinta) library.

The *X. maculatus* library was screened with a two times gel-purified fragment from the 3' kinase domain of the RSV v-*src* gene (600 bp Pst I fragment F [65], see also Appendix Fig. 1). Approximately 5×10^5 pfu total (i.e. approximately 15 genome equivalents) were plated onto ten 150mm plates and plaque-lifts made using Colony/Plaque Screen filters (NEN). The filters were pre-hybridized and hybridized in batch. Approximately 1×10^6 cpm (Cherenkov) from the nick-translated fragment was added per ml of hybridization solution (see Materials and Methods) and the filters incubated overnight. Positive clones (including both strong and weaker hybridizers) were picked and plaque-purified by two more cycles of plating, hybridization and re-picking. Twenty clones were purified, their clonal status confirmed and plate lysate stocks prepared.

Initial characterization of clones

DNA was prepared from liquid lysates of twelve of the twenty clones and restricted initially with a small number of enzymes. These initial digests sufficed however to show that several of the isolates were clearly related to other independent isolates and were in fact overlapping clones (to varying extents) from the same genomic sequences. It would appear that the twelve clones brought through this sort of very preliminary analysis are in actual fact derived from possibly six or seven different genomic sequences, and that multiple isolates have occurred from several of these.

Xiphophorus src-related clones: probable relationships based upon initial restriction enzyme digestion

19-4	20-2	22-1	20-1	17-1	17-4	22-2
20-3	17-3			18-1		
18-2				19-1		

A number of the DNA samples prepared from these twelve clones were subjected to a further series of analyses, based upon DNA sequence homology as assayed by DNA hybridization, in an attempt to define further their relationship to *src* and other members of the *src* gene family:

First, undigested DNA from nine of these twelve clones was run in duplicate gels, blotted, and then hybridized with the four Pst I fragments which cover the RSV v-*src* gene (see Appendix Fig. 1) as well as with a v-*yes* kinase domain probe. This showed that under the stringency conditions employed none of the clones assayed reacted significantly with the probe derived from

the very 5' end of the *v-src* gene - that is, the signal was identical to the cross-hybridization observed with a lambda marker fragment (Fig. 3d). In addition, several of the clones hybridized with the 350 bp *v-src* fragment (see Appendix, Fig. 1), indicating a relationship of these clones to *v-src* which extends outside of the kinase domain (eg. predominantly clones 18-2, 20-1 and 20-3 in Figure 3c). All clones tested, with the exception of clone 17-4, were reactive with the Pst I 250 bp *v-src* fragment, which contains the 5' region of the kinase domain sequence. Clones that hybridize with the 600 bp fragment (used in the library screening) but not to the other *v-src* probes are difficult to interpret following this sort of analysis because they could simply represent partial clones of a fish *src*-related gene (eg. clone 17-4). Several of the clones (eg. 17-3, 18-1 and 20-2) preferentially hybridized with the *v-yes* probe, with clone 20-2 giving the strongest signal from amongst these clones (Fig. 3e).

Second, eleven of the plate lysate phage stocks (including eight of the twelve clones above) have been plated out in an ordered array under conditions that lead to uniform and large plaques (approx. 5mm) [24a]. Plaque-lift filters from these plates have been hybridized with the four *v-src* Pst I fragments, as well as with several fragments derived from the kinase domains of other *src*-related genes. The hybridizations with the *v-src* probes (Fig. 4) show that clones 19-3, 19-4, 20-1, 20-3 and 21-1 are reactive with the two *v-src* probes covering the kinase domain (600 and 250 bp Pst I fragments) as well as to the third *v-src* probe derived from outside the kinase domain (350 bp Pst I fragment), confirming the results above for clones 20-1 and 20-3 at least. These clones would appear to be related, and in fact restriction mapping has shown for example that the entire 19-4 insert is included within clone 20-3. Hybridization with a purified kinase domain fragment from the *v-yes* gene demonstrates clearly that clone 22-1 gives the strongest hybridization signal, with clones 17-2, 20-2 and 22-2 also giving relatively strong signals (Fig. 5). Clones 19-3, 19-4, 20-1, 20-3 and 21-1 are also clearly reactive with this fragment, but to a reduced degree. When a duplicate filter to the one used with the *v-yes* probe is hybridized with a purified fragment from the mouse *c-hck* gene, clone 22-2 gives by far the strongest signal, with clone 22-1 (the *yes*-related clone, see above) giving rise to a much reduced signal (Fig. 5). These analyses have also been conducted with hybridization probes derived from the *v-fgr* and *c-hck* genes (see Figure 5). The results of these 'giant-plaque' hybridizations are summarized in Table 1.

Third, a number of duplicate slot-blot filters have been prepared that contain several viral oncogene sequences known to encode the kinase domains of tyrosine kinases (*v-src* [65], *v-yes* [157], and *v-fgr* [211], viral representatives of the *src* gene sub-family; *v-lps* [138], *v-fms* [73], *v-fes* [277], *v-ros* [212], *v-abl* [101] and *v-erb B* [333], viral representatives of non *src*-family genes encoding tyrosine kinases); as well as a serine/threonine kinase (*v-mos*) [247] and two viral genes unrelated to tyrosine kinases as negative controls (*v-Ha ras* and *v-sis*) [62,67]. Fragments that hybridize with the *v-src* Pst I 600 bp probe from clones 19-4, 22-1 and 20-2 have been gel-purified and used as hybridization probes on the slot-blot filters. The results of these hybridizations are presented in Figure 6a. The clone 19-4 fragment hybridizes predominantly with the *v-src* sequence and cross-hybridizes to some extent with the *v-fgr* sequence. The reactivity of this fragment with the *v-yes* sequence under these stringency conditions is very low (i.e. close to background). The clone 22-1 fragment hybridizes

TABLE 1 Hybridization intensity of various Xiphophorus src-related clones to different src gene-family members in 'glant plaque' analyses.^a

	600	250	350	400	YES	FGR	LCK	HCK
17-1	+	+	-	-	+/-	+	+	++
17-2	+++	+/-	-	-	++	-	-	+
17-4	+/-	+/-	-	-	-	-	-	-
19-3	+++	+++	+++	-	+	++	-	++
19-4	+++	+++	+++	-	+	++	-	++
20-1	+++	+++	+++	-	+	++	-	+
20-2	+++	+	+/-	-	+++	-	-	+
20-3	+++	+++	+++	-	+	++	-	++
21-1	+++	+++	+++	-	+	++	-	++
22-1	+++	+	-	-	++++	+/-	+	+++
22-2	+++	+/-	-	-	++	-	+++	+

^a Relative strength of the hybridization of the eleven src-related phage stocks tested hybridized with the four v-src probes and the four probes prepared from other src-family members.
 - = background hybridization; +/- = faintly detectable but definite hybridization signal; +, ++ and +++ = increasing relative hybridization signal; approximate due to comparison between different hybridizations.

most strongly to the v-yes sequence, and displays cross-reactivity at a reduced level to the v-src sequence and, further reduced, to v-fgr. The clone 20-2 fragment gives a very interesting hybridization pattern, in that the predominant hybridization is to v-src and v-fgr in that order (i.e. similar to clone 19-4) but in comparison to clone 19-4 is also cross-reactive to v-yes. In addition, the 20-2 fragment displays a weak but definite cross-reactivity to several other viral genes encoding tyrosine kinases, but which are not members of the src gene family (eg. v-fes, v-fms and v-ros).

Fourth, the hybridization signals were compared when a chicken genomic Southern blot was probed with the src-related sequences from clones 19-4 and 22-1 versus the v-src and v-yes kinase domain probes. As shown in Figure 6b, the v-yes kinase domain probe gives rise to two equally intensely hybridizing bands of approximately 2.4 and 2.0 kb (lane A) while the v-src kinase domain probe gives rise to a single hybridizing band of approximately 12-14 kb (lane B).

The clone 19-4 fragment (probe A, Figure 10a) gives the same hybridization pattern as the v-src probe. The clone 22-1 fragment hybridized strongly with one of the two fragments

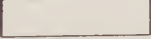
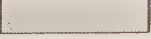
detected with the v-yes probe (the 2.4 kb fragment), and showed a very weak hybridization with the other fragment. In addition, the fragment shows some degree of cross-reactivity with the chicken c-src sequence (12-14 kb), as well as an additional sequence probably representing another closely related member of the src gene family in the chicken genome (approximately 4.4 kb; possibly c-fgr - see Figure 6a). In this sense the chicken genomic Southern blots confirm the results obtained with the slot-blot filters for clones 19-4 and 22-1.

Fifth, the purified hybridizing fragments from clones 19-4, 22-1 and 20-2 have been used as probes (under stringent hybridization and washing conditions - see Materials and Methods) on genomic Southern blots prepared from *Xiphophorus* DNA. Figure 7a and b illustrates representative results for the clone 19-4 fragment (hybridization probe A, Figure 10a). These results demonstrate: 1) the bands detected are in all cases consistent with the clone restriction maps (see below), indicating the absence of any major cloning artifacts in these clones during the preparation of the library and sub-cloning; 2) in most cases one or two bands are detected, the intensities of which are consistent with these sequences being present in the genome as unique sequences, as well as indicating the lack of repetitive sequences within this 1.3 kb fragment used as the probe; 3) some of the bands detected in genomic Southern blots using viral probes under moderate stringency (see for example Fig. 2a) can be accounted for by clones 19-4, 22-1 and 20-2, but not all, suggesting that additional src-related sequences exist in the fish genome - this result is summarized in Table II; and 4) the hybridization probe, a 1.3 kb Pst I fragment from a *X. maculatus* genomic clone, also detects, although to a slightly lesser extent, an approximately 1.3 kb fragment in Pst I digested *X. helleri* DNA (Figure 7b).

Lastly, clones 19-4, 22-1 and 20-2 (amongst others) were subjected to some preliminary restriction enzyme mapping and hybridization with v-src and related probes. Figure 8 presents representative results for several single and double digests of the lambda clone 19-4. Figure 8A shows that specific fragments of clone 19-4 are strongly reactive with the two kinase domain probes (Pst I 600 bp and 250 bp fragments) as well as a v-src probe derived from outside the kinase domain (Pst I 350 bp fragment). Figure 8B shows that clone 19-4 is not only reactive with a v-src kinase domain probe (the Pst I 250 bp fragment) but that under very reduced stringency conditions can give rise to very weak signals with a probe from the divergent extreme 5'-end of the viral gene (Pst I 400 bp fragment). Figure 9 presents the preliminary restriction maps for these three clones and several others (20-1 and 22-2), as well as a summary of some of the hybridization data with the v-src probes and several other src family probes. This figure serves to highlight the independent nature of the clones (i.e. they share no apparent relationship in terms of their restriction maps or their hybridizing fragments). Based upon the analyses presented above it was clear that clone 19-4 (and related overlapping clones) contained the most src-related sequence and was, therefore, subjected to further analysis. Restriction enzyme digestions showed that clone 19-4 did not contain an internal Eco RI site, and the entire Eco RI insert of 19-4 was therefore subcloned into pUC 9. Further mapping, subcloning and sequencing were performed with this plasmid or subclones derived from it. To aid in mapping and subcloning of such a large genomic fragment, the 19-4 Eco RI insert was purified free of plasmid vector sequences (see Materials and Methods) and this purified fragment used in certain mapping and subcloning applications.

TABLE II: SRC-RELATED SEQUENCES IN GENOMIC SOUTHERN BLOTS^a

Eco RI	Bam HI	Hind III
22.0 +++	23.0 ++	30.0 +
13.5 ++	15.5 +/-	23.0 +/-
9.8 +/-	11.0 +	11.0 +
7.5 +	9.6 +/-	8.5 +/-
6.9 +	8.4 +++	8.3 ++
5.4 +++	7.8 +	7.0 ++
4.2 ++	7.4 +	6.3 +
2.0 +/-	6.8 +/-	6.0 +
	6.0 +	5.0 +/-
	4.5 +++	4.7 +
		4.2 +/-
		2.9 +/-
		2.5 +
		2.0 +
		1.8 +

 Xsrc 19-4  Xsrc 22-1  Xsrc 20-2

^a The *Xiphophorus maculatus* fragments that hybridize with the v-src 600 bp kinase domain probe under standard stringency conditions (see Materials and Methods) following digestion with Eco RI, Bam HI or Hind III are presented, including the relative strength of the hybridization signal for each band. Superimposed are the bands detected by each of clones 19-4, 22-1 and 20-2 under stringent conditions.

Presented below is the full kinase domain sequence for the *X. maculatus* *src*-related clone 19-4. For the purposes of comparison, the available kinase domain sequence from the *yes*-related clone 22-1 has been utilized. The further mapping (i.e. beyond that shown in Figure 9), sub-cloning and sequencing of clone 22-1 has been performed by G. Hannig within our laboratory, as part of his ongoing Ph.D thesis.

Restriction fragments of clone 19-4 that hybridize with the viral kinase domain probes have been subcloned and sequenced. Figure 10a presents a restriction map of clone 19-4, including the sequencing strategy, and demarcation of the regions identified to date demonstrating sequence relationship to the *v-src* kinase domain sequences. Also included in Figure 10a are the hybridization probes from clone 19-4 that have been used during various aspects of this work. Figure 10b presents the nucleic acid sequence obtained from this clone, including the theoretical translation. The presumptive coding regions of clones 19-4 and 22-1 are very closely related at both the nucleic acid and at the amino acid level (74% and 87% identity respectively), whereby most base substitutions are in the third or 'wobble' position of the codon (87 of 115; 75.6%) and most amino acid substitutions are conservative - i.e. of similar size and/or chemical nature (11 of 18; 61%; Robertson, S.M. *et al.* submitted).

By virtue of nucleic acid sequence comparison of the fish clone 19-4 to the chicken *c-src* [305] and the human *c-src* [16] and *c-fgr* [153,217] genes, presumptive exons could be identified, which has since been duplicated for clone 22-1 (G. Hannig, unpublished; Robertson, S.M. *et al.* submitted; see also section on gene evolution). Figure 11 presents a direct comparison of part of the genomic sequence of clone 19-4 (600 bp Pst I/Xba I fragment; Probe B, Figure 10a) to the published chicken *c-src* sequence (with introns deleted). The diagonal lines signify a high level of sequence similarity between the two sequences and the horizontal breaks from the diagonal signify stretches of non-similarity (i.e. potential introns) present in the clone 19-4 sequence. Specifically, the horizontal breaks in Figure 11 represent (from bottom to top) the introns following the fish homologues of exons 8, 9 and 10 (numbered according to the chicken *c-src* gene). The degree of shift from the diagonal resulting from such a break is also a measure of the length of the introns in the fish gene, and it can be seen that introns 8 and 9 (like the homologous introns in the chicken gene) are relatively small- 101 and 104 basepairs respectively, compared to 78 and 61 basepairs in the chicken gene [305]. However, in the rest of the gene there is a high degree of variability in the intron sizes when the fish gene is compared to either the human *c-src* gene or the chicken gene (Figure 12b). For example, the intron following exon 7 is approximately 2300 bp in human, 85 bp in the chicken and approximately 450 bp in the fish *c-src* gene. Likewise, the intron following exon 10 is 160bp, 118 bp and 790 bp in the human, chicken and fish genes respectively. By comparison of the genomic sequences the probable exon/intron boundaries of the fish *c-src* gene could be precisely located. All introns followed the GT/AG rule [207] for splice donor and acceptor sites (Figure 12a). The location of the presumptive introns in the fish *c-src* gene matched exactly the location of the homologous introns in the chicken and human *c-src* genes as well as the human *c-fgr* gene. Figure 13 compares the theoretical exon/intron configuration of the fish clone 19-4 to the available data concerning the

entire *src* gene family, which clearly demonstrates that the genomic configuration for this gene is identical to the arrangement in higher vertebrates and mammals. An open reading-frame covering the entire sequenced region resulted upon removal of these presumptive introns from the primary sequence of clone 19-4.

The *Xiphophorus* library was screened with the express intent of isolating several members of the fish *src* gene family. Being a more sensitive measure of gene relationships, the kinase domain nucleic acid and amino acid sequences of clone 19-4 and the partial sequences from clone 22-1 (G. Hannig, unpublished) were therefore compared to all known members of the *src* gene family as well as members of several other tyrosine kinase sub-families. As shown in Table III, both sequence comparisons confirm clone 19-4 as being most closely related to *c-src*, and at the amino acid level clone 22-1 as being most closely related to *c-yes*. Clones 19-4 and 22-1 will therefore hereafter be referred to as *Xiphophorus maculatus c-src* (*Xsrc*) and *c-yes* (*Xyes*) respectively.

Figure 14 compares the translation of *Xsrc* to the amino acid sequence of the homologous human gene, SRC [16]. There exists amino acid sequence data for a large number of protein kinases from a variety of organisms, and comparison of these sequences has led to the definition of certain sequence motifs diagnostic for protein kinases. The most definitive of these are the motif G-G-G-V...⁽¹⁵⁾...A-K of the putative ATP-binding site (residues 277-298 of SRC), and the motif (H/Y)RDL...⁽¹⁶⁾...DFG.....APE (residues 388-435 of SRC)[19a,144]. *Xsrc* contains these motifs, and in exactly the same relative spacing as in the human SRC gene. In addition, there are a number of amino acid residues which, although not as diagnostic as the two preceding motifs, are nonetheless highly conserved amongst protein kinases in general. These residues are marked by open circles in Figure 14. The *Xsrc* translation contains 60 out of 60 of these residues. There are 24 residues within the kinase domain that are shared by all known (n 0 27 [124a]) tyrosine kinases (marked by closed circles in Figure 14). The *Xsrc* translation contains all 24 of these residues. In conjunction with the second sequence motif presented above, there are also two short sequence motifs that distinguish a tyrosine-specific protein kinase from one with serine/threonine specificity - (H/Y)RDLBAAN or (H/Y)RDLAARN is diagnostic for a tyrosine kinase whereas (H/Y)RDLKPEN is more commonly found in the serine/threonine protein kinases; (E/L)E(I/V)(K/R)W(I/M)APE is diagnostic for a tyrosine kinase, whereas (G/V)(I/S)XX(V/E)XAPE is diagnostic for a serine/threonine-specific protein kinase. The *Xsrc* translation contains both sequence motifs diagnostic for tyrosine-specific protein kinases (marked as TK I and TK II in Figure 14).

The regulation of the enzymatic activity of the viral and cellular *src* proteins of the chicken has been investigated in some detail in the hope of, firstly, helping to define the normal cellular function of pp60^{C-Src} and, secondly, in attempting to define the transformation-associated functions of pp60^{V-Src}. These studies indicate that phosphorylation of pp60^{C-Src}, particularly of a tyrosine residue near the C-terminus (tyr-527 of the chicken pp60^{C-Src}) plays a critical role in the regulation of *src*-derived tyrosine kinase activity in normal cells [for a review see 6]. The fish *Xsrc*, as well as the human SRC and YES translations contain this residue at the analogous position, raising the possibility that the fish protein (as well

as the human SRC and YES proteins) are regulated in a similar fashion to the cellular *src* protein of chicken. In addition, the *v-src* studies have highlighted the role of phosphorylation of another tyrosine residue (tyr-416 of chicken pp60^{c-src}; included in the motif RDL...DFG...Y...APE) in regulation of the viral transforming tyrosine kinase activity. The fish *c-src* gene (as well as SRC and YES) encodes the same residue at the same position within this precisely maintained sequence motif.

TABLE III:

Sequence comparison of f1ish src-related clones 19-4 and 22-1
to src gene-family members^a

	<u>AMINO ACIDS</u>		<u>NUCLEIC ACIDS</u>	
	<u>19-4</u>	<u>22-1</u>	<u>19-4</u>	<u>22-1</u>
HUMAN C-SRC	91.4	85.1	79.6	76.9
HUMAN C-YES	84.4	90.5	70.5	76.1
HUMAN C-FYN	75.4	78.9	67.7	72.4
HUMAN C-FGR	73.7	72.3	69.8	70.7
HUMAN C-HCK	65.9	68.3	65.7	67.2
HUMAN C-LYN	65.1	63.3	65.1	63.8
MOUSE C-LCK	63.4	66.7	63.2	65.9
CHICKEN C-TKL	64.2	69.2	65.7	67.5
HUMAN C-ABL	62.5	59.4	58.9	55.2
HUMAN C-FPS/FES	40.9	39.1	50.7	50.1

^a Comparison of nucleic acid and amino acid identity between the sequences determined for Xiphophorus clones 19-4 (exons 7-12) and 22-1 (exons 7, 10 and 11 - G. Hannig, unpublished; Robertson, S. M. et al, submitted for publication) to the sequences published for the src gene-family members characterized to date. Also included are two human sequences (c-abl and c-fes/fps) which encode tyrosine kinases but which are not members of the src gene family.

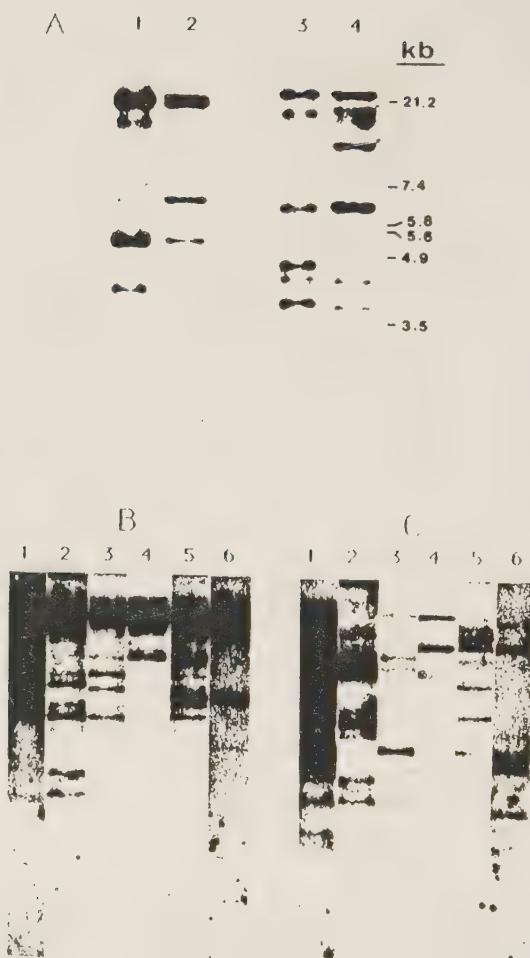


FIGURE 2

Family of src-related sequences detected in the *Xiphophorus* genome.

The two viral hybridization probes used: the 600 bp v-src Pst I fragment derived from the kinase domain of Rous sarcoma virus, and the 1.1 kb v-yes Pst I fragment containing the kinase domain of avian Y73 virus.

A) Hybridization of *X. maculatus* DNA (lanes 1 and 2) and *X. helleri* DNA (lanes 3 and 4) with the v-src kinase domain probe (lanes 1 and 3) and the v-yes kinase domain probe (lanes 2 and 4). Standard hybridization/washing stringency conditions as described (Materials and Methods).

B) and C) Hybridization of *Xiphophorus* DNA with src-family members.

Xiphophorus andersi DNA digested with Eco RI (B) and Hind III (C) hybridized with gel purified fragments derived from the kinase domains of the following genes: v-src (lane 1), v-yes (lane 3), v-lgr from Gardner-Rasheed feline sarcoma virus (lane 4), c-lck from the mouse LSTRA cell line (lane 5), and human c-lck (lane 6). In addition, the v-src 350 bp Pst I fragment, which derives from outside the v-src kinase domain, is included (lane 2).

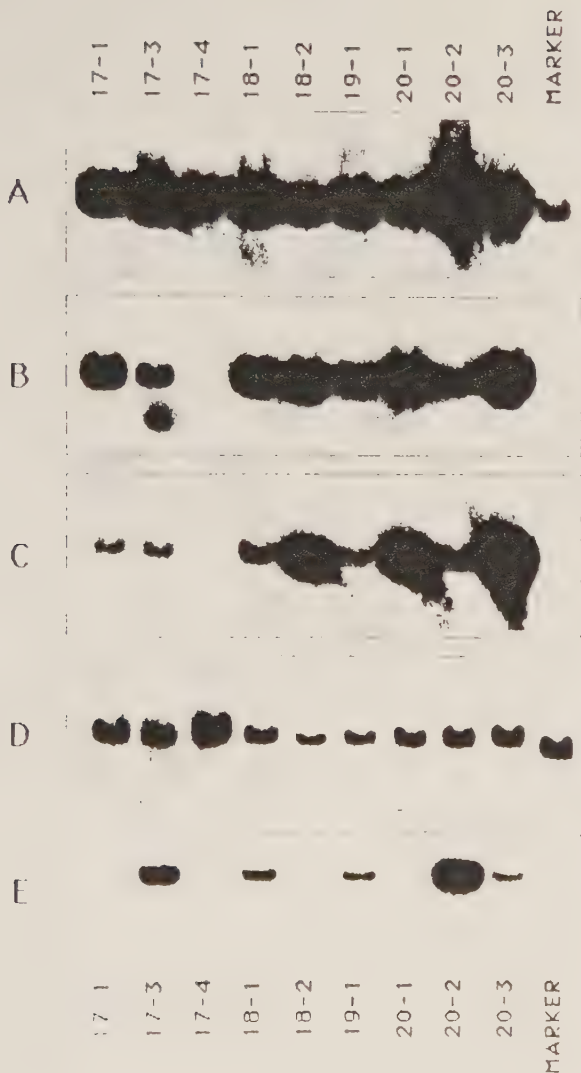


FIGURE 3

Hybridization of purified lambda DNA with the four *v-src* probes and the *v-yes* probe.

Purified lambda DNA (undigested) from nine *src*-related clones was run in a 0.8% agarose gel, transferred to a nylon membrane, and hybridized with: A) *v-src* 600 bp *Pst* I fragment (kinase domain); B) *v-src* 250 bp fragment (kinase domain); C) *v-src* 350 bp *Pst* I fragment; D) *v-src* 400 bp *Pst* I fragment (5' end of *v-src* gene), and E) the 1.1 kb *Pst* I *v-yes* fragment (kinase domain). Standard stringency conditions in A), B), C) and E); reduced stringency conditions in D) (see Materials and Methods)

Note: D) Includes a lane with marker DNA (lambda/*Hind* III). Under the reduced stringency conditions employed the probe reacts with a lambda fragment to an equivalent extent as with the nine *src*-related clones.

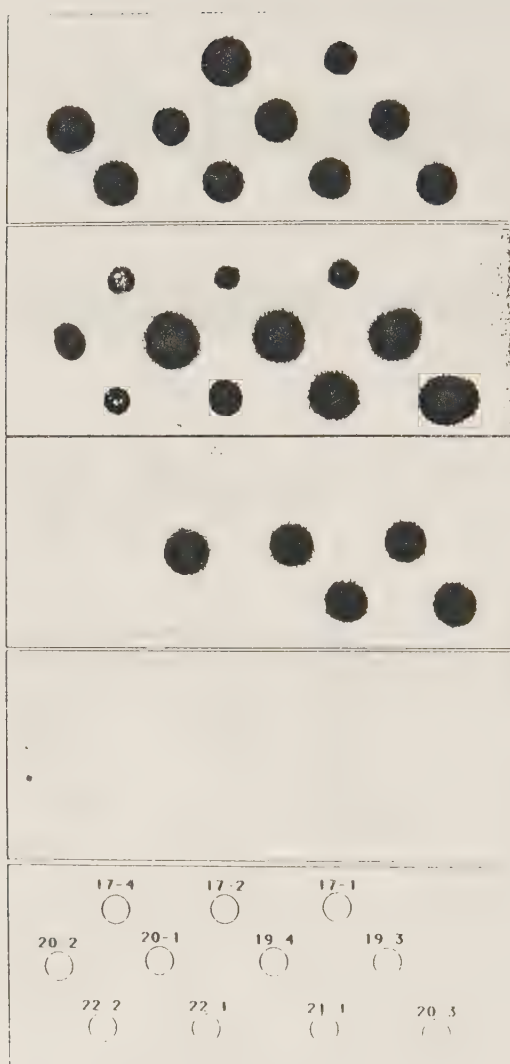


FIGURE 4

Giant plaque hybridizations with the four *v-src* Pst I probes.

Eleven *src*-related phage stocks were plated out in an ordered array (see Materials and Methods) and the resulting large plaques transferred to nylon membranes and hybridized with the four *v-src* Pst I fragments. Hybridizations with: A) *v-src* 600, B) *v-src* 250, C) *v-src* 350, and D) *v-src* 400. Standard stringency conditions were employed (Materials and Methods).

Xiphophorus src related clones

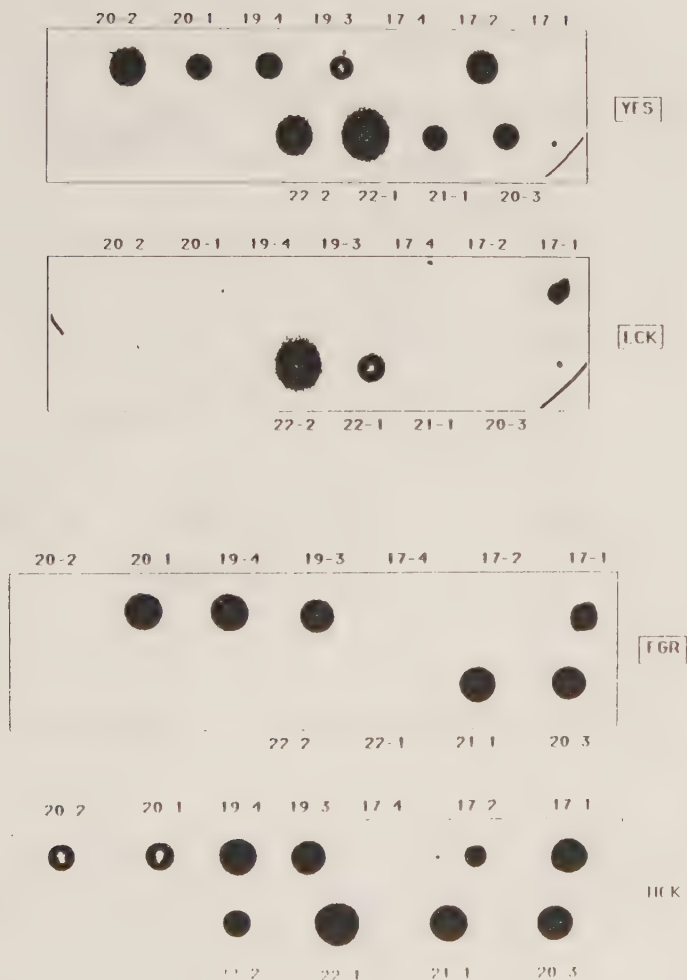


FIGURE 5

Giant plaque hybridizations with probes derived from various *src*-family members.

An ordered array of large plaques from eleven of the *src*-related clones transferred to nylon membranes and hybridized under standard conditions (see Materials and Methods) to kinase domain probes prepared from: *v-src*, *c-lck*, *v-lgr* and *c-hck*.

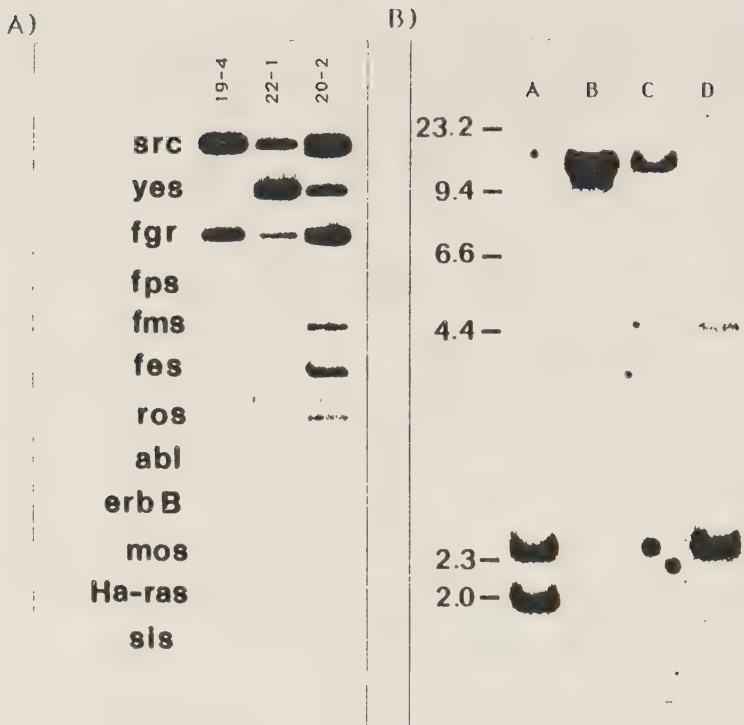


FIGURE 6

Relationship of *src*-related clones (19-4, 22-1 and 20-2) to tyrosine kinase family members.

A) Slot blot containing equal molar amounts of DNA from nine viral oncogenes of the tyrosine kinase gene super-family (*v-src*, *v-yes*, *v-fgr*, *v-fps*, *v-fms*, *v-fes*, *v-ros*, *v-abl* and *v-erb B*), one viral member of the ser/thr class of protein kinases (*v-mos*), and two viral oncogenes not related to the protein kinases (*v-Ha-ras* and *v-sis*) as negative controls. The hybridizing fragments from the three *Xiphophorus* *src*-related clones 19-4 (Pst I 1.2 kb; probe C, Fig. 10a), 22-1 (Hind III 3.1 kb) and 20-2 (Bam HI 4.6 kb) were gel-purified and used as hybridization probes onto duplicate slot blot filters. Standard stringency conditions (Materials and Methods) were employed.

B) Chicken genomic DNA digested with Eco RI and hybridized with: the *v-yes* 1.1 kb kinase domain fragment (lane A); the *v-src* 600bp kinase domain probe (lane B); the *Xiphophorus* clone 19-4 Pst I 1.2 kb fragment (lane C), and the *Xiphophorus* clone 22-1 3.1 kb fragment (lane D).

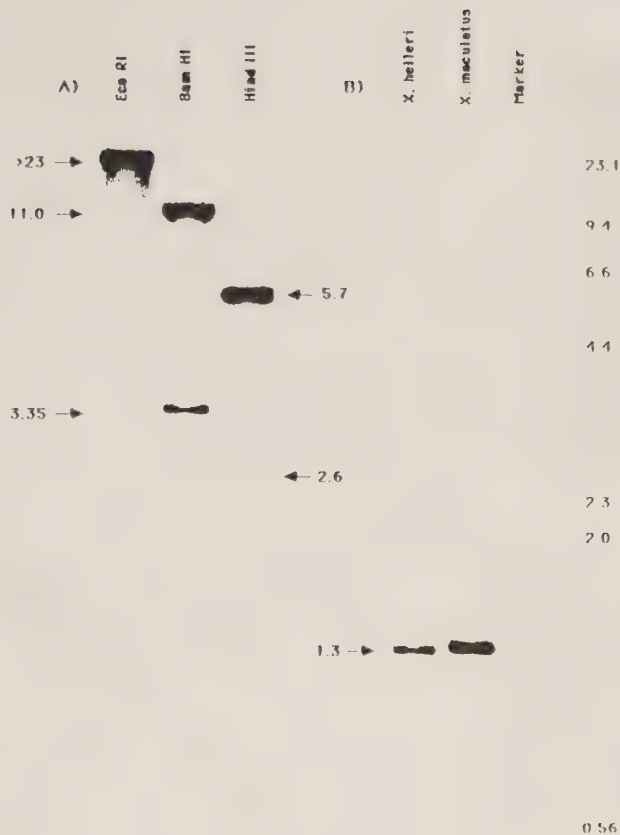


FIGURE 7

Xiphophorus genomic Southern blots hybridized with *src*-related clone 19-4.

A) *Xiphophorus maculatus* DNA digested with the marked restriction enzyme and hybridized under stringent conditions (see Materials and Methods) with the clone 19-4 Pst I 1.3 kb fragment (probe A, Fig. 10a). B) *X. helleri* and *X. maculatus* DNA digested with Pst I and hybridized under stringent conditions with the *X. maculatus* clone 19-4 Pst I 1.3 kb fragment. The arrow marks the 1.3 kb band detected in both samples.

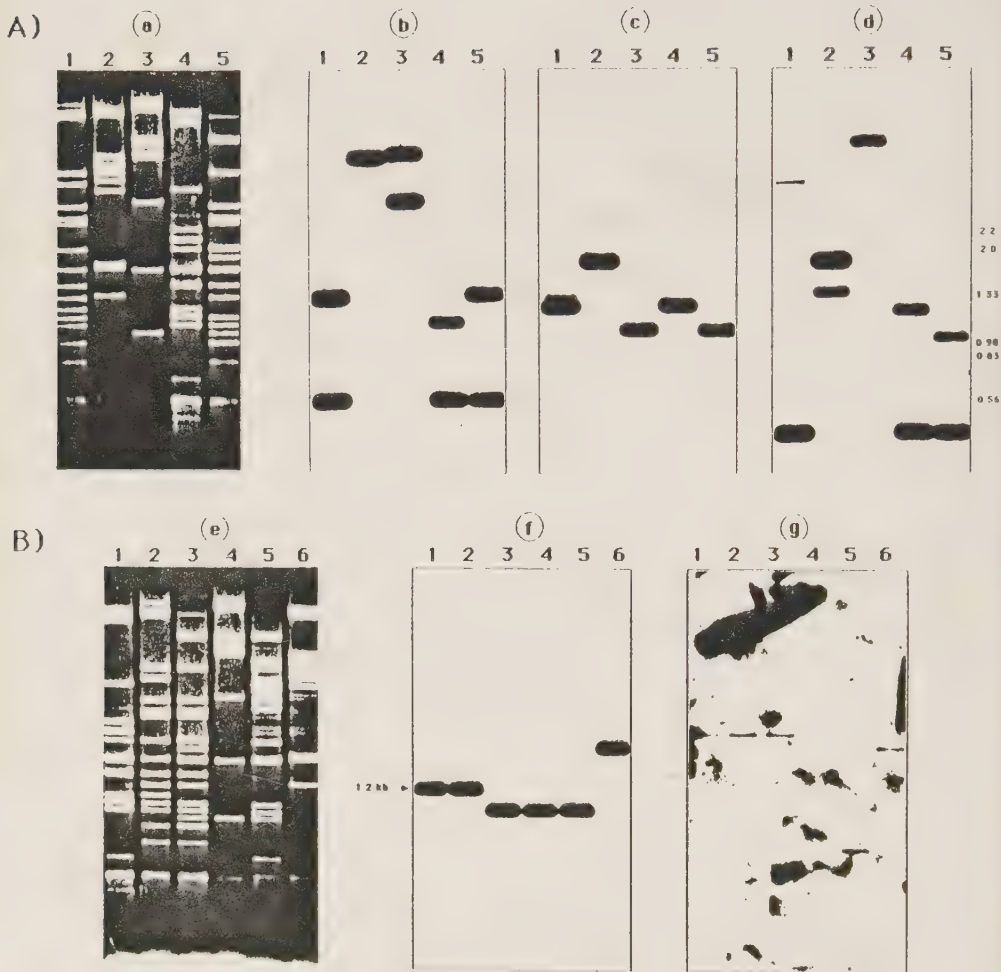


FIGURE 8

Restriction enzyme mapping/hybridization of clone 19-4.

b) - d): Identical filter of lambda DNA from clone 19-4 digested with various enzymes used in restriction mapping and hybridized successively with: b) the *v-src* 600 bp Pst I fragment; c) the *v-src* 250 bp Pst I fragment and d) the *v-src* 350 bp fragment. a) The ethidium-bromide stained gel for b) - d).
f) and g): Identical filter of lambda DNA from clone 19-4 digested with various enzymes used in restriction mapping and hybridized successively with: f) the *v-src* 250 bp Pst I fragment, and g) the *v-src* 400 bp Pst I fragment. e) The ethidium-bromide stained gel for f) and g). In b) - d) and f) standard stringency, and in g) reduced stringency conditions were used (see Materials and Methods).

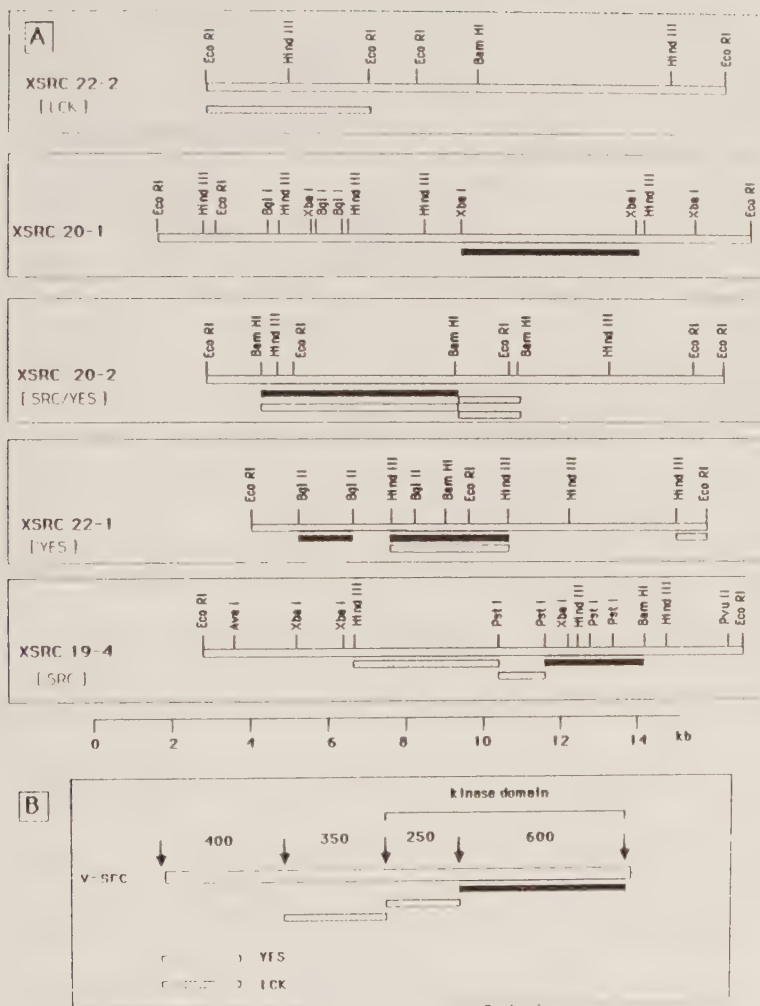


FIGURE 9

Restriction enzyme maps of *Xiphophorus src*-related clones 22-2, 20-1, 20-2, 22-1 and 19-4.

A) Restriction maps of the lambda clone plus representative hybridization results using various v-src and src-family probes. B) Hybridization probes utilized in A).

FIGURE 10a: Restriction map and sequencing strategy for clone 19-4. Exons marked by rectangles.

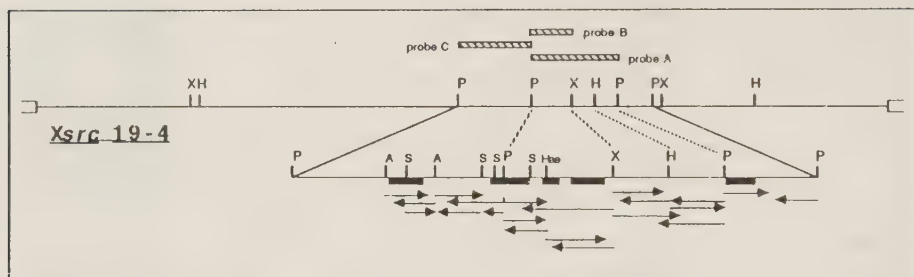


FIGURE 10b: Nucleotide sequence and theoretical translation for kinase domain of clone 19-4. Exon/intron boundaries marked by asterisks.

* (7)

HisAlaAspGlyLeuCysHisThrLeuThrAspValCysProThrLeuLysProGlnThrGlnGlyLeuSerLys
agcacgcgatgggctgtgtcacactctcagcgacgtttgtcccactctgaagcctcagactcagggttatccaag

AspAlaTrpGluIleProArgGluSerLeuArgLeuGluValLysMetGlyGlnGlyCysPheGlyGluValTrpArg
gatgcctgggagatccgagagagtcacttcgcctggaggtcaagatgggacagggtcgttcggagaggtctggaga

* (8)

GlyThrTrpAsnGlyThrThrGlnValAlaIleLysThrLeuLysProGlyThrMetSerProGluAlaPheLeuGln
ggaacatggaacggcaccacacaggtggcgatcaagacgctgaagccggcaccatgtccccgaggcggttcctgcag

GluAlaGlnValMetLysLysLeuArgHisGluLysLeuValGlnLeuTyrAlaValValSerGluGluProIleTyr
gaagctcaggtcatgaagaaactgagacacgagaagctggttcagctgtacgccgtggttctcaggagaccaatctat

* (9)

IleValThrGluPheMetAspGlnGlySerLeuLeuGluPheLeuLysGlyGluThrValArgTyrLeuArgLeuPro
atcgtaacagagttcatggaccaaggcagcttatggagttctgaaggcgagacagtcgcatatctacgcctccct

* (10)

GlnLeuValAspPheAlaSerGlnIleAlaSerGlyMetAlaTyrValGluArgMetAsnTyrValHisArgAspLeu
cagctggtggactttgcctctcagattgcttcaggatggcctacgtagagaggtgaattacgttcatagagacctc

ArgAlaAlaAsnIleLeuValGlyAspAsnLeuValCysLysValAlaAspPheGlyLeuAlaArgLeuIleGluAsp
cgagccgctaacatcctgggtggagataatctggttgcaagtggtgattttggtttggctcgctgattgaggtat

* (11)

AsnGluTyrThrAlaArgGlnGlyAlaLysPheProIleLysTrpThrAlaProGluAlaAlaLeuTyrGlyArgPhe
aacgaatatcacgccaggcaaggcgccaagttcccatcaagtggaaggctcctgaggcggcgtctacggacgattc

(12) *

ThrIleLysSerAspValTrpSerPheGlyIleLeuLeuValGluLeuAlaThrLysGlyArgValProTyrProGly
accattaaatctgacgtctggtcatttggatcctgctggttgaactggccacaaaggtcgctgcccctaccaggc

MetValAsnArgGluValLeuAspGlnValGluArgGlyTyrArgMetArgCysProSerGluCysProPheSerLeu
atggtgaacagagaggttctggatcaggtggagagggatcacaggtgcggtgtccgtctgaatgtcctttctctctc

HisGluLeuMetLeuAsnCysTrpArgLysGluAlaGluGluArgProThrPheGluTyrLeuGlnAlaPheLeuGlu
catgaactgatgctgaactgctggaggaaggagggcagagagggccaccttcaggtacctgcaggccttctctggag

AspTyrPheThrSerThrGluProGlnTyrGlnProGlyAspAsnLeuSTOP
gattacttcacctccactgagcctcagtcaccagcctggggataacctgtaa

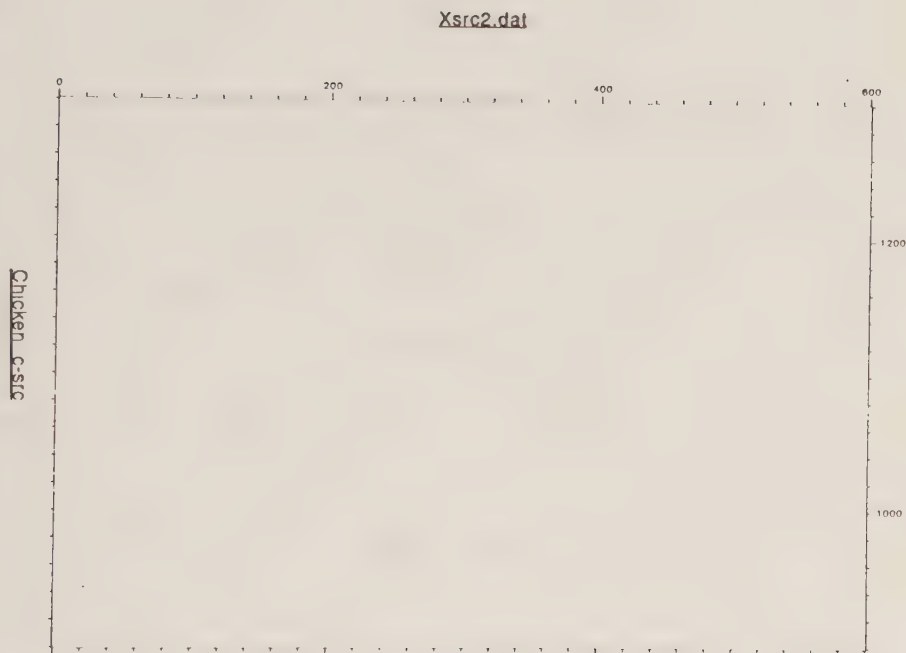
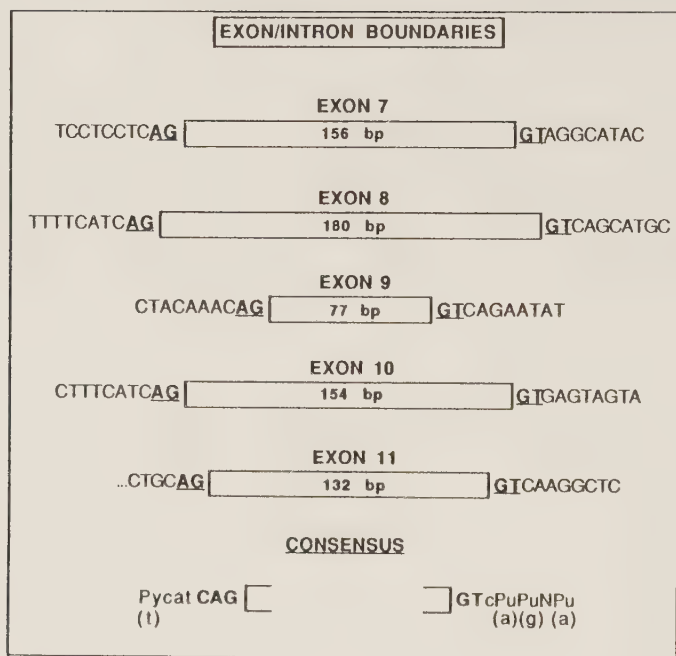


FIGURE 11

Dot plot comparison of partial *Xsrc* sequence to a portion of the chicken *c-src* gene (introns removed)

The program Compare of the University of Wisconsin Genetics Computer Group sequence analysis program package was used to compare portions of the fish and chicken *c-src* genes. A moving window of 21 bases compares in every possible register the two sequences and a dot results where at least 14 bases match within the window. Strings of dots along the diagonal indicate sequence similarity. Conversely, breaks in dot-strings denote regions of relative non-similarity - i.e. potential introns within the *Xsrc* genomic sequence.

A)



B)

		INTRON SIZES									
		2	3	4	5	6	7	8	9	10	11
HUMAN	C-SRC						2300	1600	780	160	280
CHICKEN	C-SRC	50	2040	390	1010	350	85	78	61	118	79
XIPH.	C-SRC					500	450	101	104	790	400

FIGURE 12

A) Exon/intron boundaries surrounding exons 7-11 (numbered according to the chicken *c-src* gene) of clone 19-4. For each boundary the GT of the splice donor site and the AG of the splice acceptor site are underlined. A consensus splice donor and splice acceptor site sequence for the fish gene is presented. B) Comparison of homologous intron sizes based upon available genomic sequences from the fish, chicken and human *c-src* genes. Intron sizes are given in basepairs.

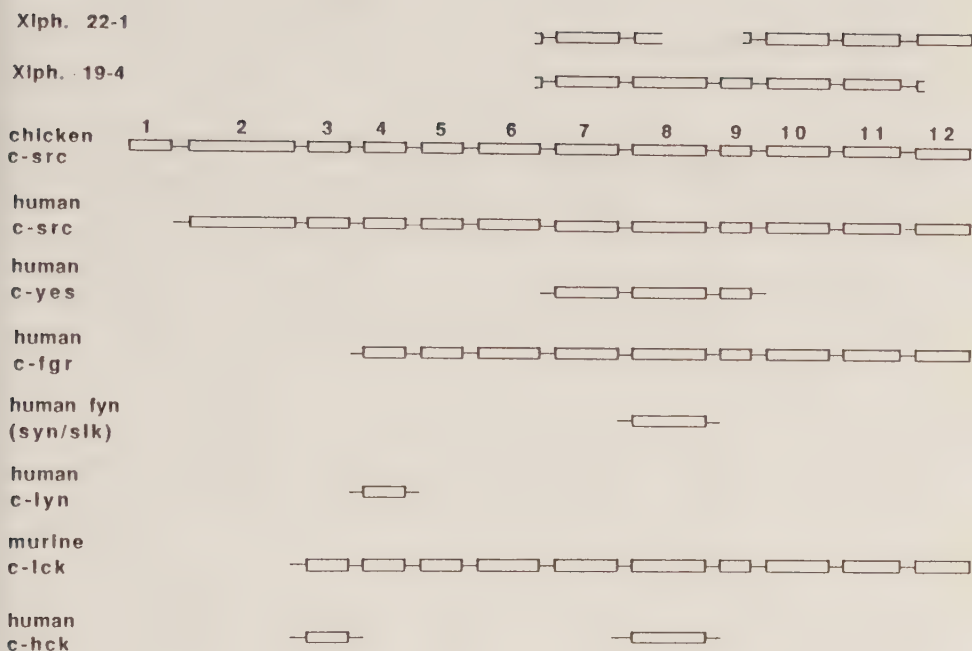


FIGURE 13

Exon/Intron configuration of the vertebrate *src* gene family.

The derived exon/intron configuration of clone 19 is compared to the available data concerning other vertebrate *src* gene-family members. Introns are not presented to scale.

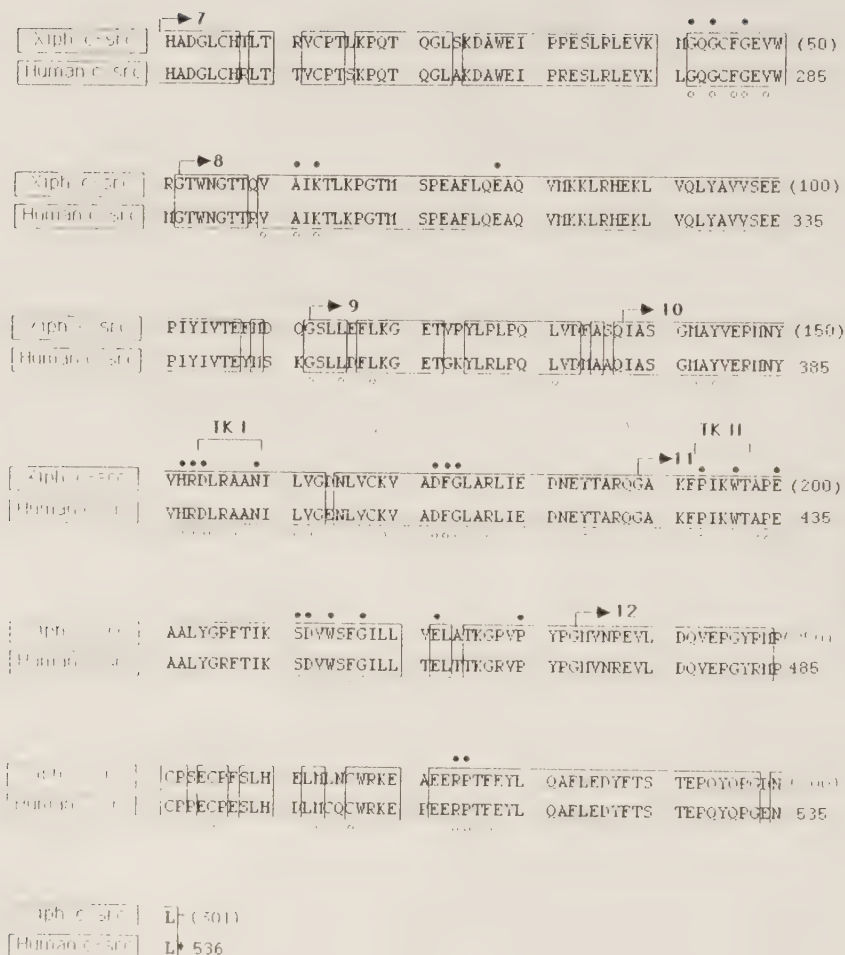


FIGURE 14

Amino acid sequence comparison of the fish clone 19-4 (Xsrc) to the human SRC gene. Identical residues are boxed. Open circles denote residues (n = 60) that are highly conserved in all protein kinases [144]. Closed circles mark residues (n = 24) that are conserved in all (n = 27) known tyrosine kinases [124a]. TK I and TK II mark two short sequence motifs diagnostic for tyrosine-specific protein kinases. Arrows demarcate exon/intron boundaries. Numbering for the human SRC protein is according to published sequences [16, 307]. Numbering for the fish protein begins from the first codon fully contained within exon 7.

III. RESULTS

2. INVESTIGATIONS OF THE ROLE OF THE C-SRC GENE IN THE CROSSING CONDITIONED MELANOMAS OF *XIPHOPHORUS*

A) NONIDENTITY OF *Xsrc* AND *mdl*

Given the correlation between the degree of deregulation of *mdl* and the level of expression of the *Xiphophorus c-src* gene [263], the question arose whether the *c-src* gene is in fact the *mdl* locus, or alternatively, whether the two loci are co-regulated due to a physical linkage to the same region of a chromosome. In order to directly test these possibilities it was first necessary to be able to differentiate the *X. maculatus c-src* gene from that of *X. helleri*. Then one would be able to assay DNA from a hybrid fish that has been repeatedly backcrossed to the *X. helleri* parent, and therefore contains only the *X. maculatus* X-chromosome (containing the deregulated *mdl* locus) effectively isolated in the *X. helleri* genetic background. The presence of the *X. maculatus c-src* gene in such fish would demonstrate linkage of the *c-src* gene to the X chromosome, and therefore linkage to the *mdl* locus. In addition, DNA was also assayed from hybrid fish where a mutant *X. maculatus* X-chromosome (see Materials and Methods; Experimental animals) was present in the *X. helleri* background. These mutant chromosomes involve translocations or deletions in and/or around the *mdl* locus, and it was hoped that if indeed the *c-src* gene did map to the *X. maculatus* X chromosome, that these chromosomes would aid in the relative physical mapping of the two loci.

Xiphophorus maculatus and *Xiphophorus helleri* brain DNAs were digested with eighteen different restriction enzymes (including both 6-base and 4-base cutters), run side-by-side on agarose gels, transferred to Gene Screen filters, and hybridized with a purified fragment from the *Xsrc* gene (the approximately 600 bp Pst I/Xba I fragment containing most of exon 8 and all of exons 9 and 10 plus intervening introns; Probe B, Fig. 10a). This probe under moderate stringency conditions detects both the *X. maculatus* as well as the *X. helleri c-src* sequences, but under very high stringency conditions almost exclusively the *X. maculatus* sequence. Under moderate stringency conditions most of the restriction digests gave rise to hybridizing bands that were identical in the two DNA samples. However, four enzymes (Acc I, Xba I, Sau 3A and Rsa I) gave rise to restriction fragment length polymorphisms (RFLP's) between the *X. maculatus* and the *X. helleri* DNA samples (data not shown). Because the Sau 3A RFLP was very readily distinguishable it was chosen as the enzyme with which to assay the various hybrid fish DNAs. The results of this analysis are shown in Figure 15. The strongly hybridizing 1.3 kb Sau 3A fragment is only detected in DNA extracted from *X. maculatus*, and is not detected in the *X. maculatus* x *X. helleri* backcross fish (Figure 15a). Figure 15b shows the result of an identical experiment whereby the stringency of the hybridization and the washing was reduced in order to more readily detect the hybridizing bands deriving from the *X. helleri* genome (which, due to the very high stringency conditions, are only faintly detectable in Figure 15a). Under the stringency conditions used in Figure 15a the *X. maculatus* sequence can be effectively differentiated from the *X. helleri* sequence. The RFLP is apparently the result of the *X. helleri* DNA containing a Sau 3A

enzyme recognition site within the hybridizing sequence that is not found in the *X. maculatus* sequence. This is indicated by the fact that the size of the two *X. helleri* bands (0.75 kb and 0.56 kb) sum together to the size of the *X. maculatus* band (Figure 15b). These data indicate that *X. maculatus* / *X. helleri* backcross hybrids with a deregulated *mdl* locus from *X. maculatus* do not contain the *c-src* gene from the *X. maculatus* parent.

B) METHYLATION STATUS OF THE *Xsrc* GENE IN RELATION TO MELANOMA FORMATION

Because of its proposed role in the regulation of gene expression (see Introduction)[49], the methylation status of the fish *c-src* gene was looked at in relation to the tumor formation.

As a preliminary investigation, testes DNA from *X. maculatus* was digested with two different pairs of restriction enzymes and then hybridized with either a *v-src*, a *Xsrc* or a *v-yes* specific fragment. The two enzyme pairs used were: 1) Dpn I and Sau 3A, which recognize the same sequence (GATC) and which Dpn I cuts only when the adenosine is methylated, but which Sau 3A cuts irrespective of its methylation status [213]; and 2) Hpa II and Msp I, which recognize the same sequence (CCGG), which Msp I cuts irrespective of methylation status of the internal C but which Hpa II does not cut when the internal C is methylated [213] (>95% of the ^mC found in eukaryotes occurs in CG doublets) [28]. Figure 16 presents the results of these initial methylation studies. It can be seen that no *v-src*, *Xsrc* or *v-yes* related sequences from the kinase domain are methylated at the adenosine site of the Dpn I/Sau 3A recognition sequence (compare lanes 1 and 2 in figure 16a, b and c - there is no suggestion of digestion by Dpn I). Second, the *v-src* and *v-yes* hybridizations show that the vast majority of *src* gene family sequences which are detectable with these probes are predominantly methylated at the CCGG sites within the testes DNA (compare lanes 3 and 4 in Fig. 16a and 16c - note the high molecular weight of the hybridization in lane 3). Third, the sequence detected with the *Xsrc* probe is flanked by two Msp I sites that are methylated at the internal cytosine residue - i.e. C^mCCGG; Hpa II resistant (compare lanes 3 and 4, figure 16b). In addition, because the Msp/Hpa recognition sequence is only four basepairs, which would be expected statistically to occur very often in the DNA (i.e. approximately once every 250 bp on average), but the hybridizing band detected in the Hpa II digests with the *v-src* and *v-yes* probes under standard stringency conditions is of high molecular weight (lane 3, Fig. 16a and c), this would suggest that the majority of Hpa II sites are methylated. The ethidium-bromide stained gel is consistent with this (Fig. 16d).

DNA was prepared from tumor-bearing hybrid fish, with melanoma, liver and brain DNA being prepared from the same group of fish so that one could compare the methylation status of the *c-src* gene in the tumor and in normal tissues from the same fish. The tumors were prepared from the approximately 25% of the backcross fish with malignant melanoma, with care taken to select only highly malignant tumors. The tumors from approximately ten fish were pooled for the DNA extraction. This DNA was digested with Hpa II and then hybridized with the *Xsrc* fragment (Fig. 17a). The parallel Msp I digestions failed due to inactive enzyme, and could not be repeated since the same DNA samples were also used in the first experiment presented in the next section, which exhausted the limited supply of this tumor DNA sample. Figure 17c

presents the results of this hybridization with the Hpa II digests. In comparison to the similar signals present in the liver and brain samples (lanes 1 and 3 respectively), the tumor DNA sample (lane 2) shows a very strong hybridizing band of lower molecular weight. Figure 17d shows the result of a hybridization of the identical filter with a fragment purified from a *X. maculatus* clone demonstrating sequence homology to several *Drosophila* homeobox-containing genes (Fig. 17b; see also Appendix 5 for characterization of this probe). It was presumed that the *Xhomeobox* probe could serve as a control for a gene not implicated in the tumor formation. As seen in Figure 17d, the Hpa II hybridization patterns are identical between the three DNA samples. In addition, a comparison of the relative hybridization signal intensities between lanes 1, 2 and 3 in parts c and d of Figure 17 is consistent with an enhanced c-*src* signal specifically within the tumor DNA sample (see below).

This experiment was repeated with a tumor DNA sample prepared from predominantly benign tumor material. These results (data not shown) are consistent with the segment of the c-*src* gene detected by the hybridization probe (i.e. exons 8, 9 and 10) being highly methylated at CCGG sequences, but do not suggest any demethylation of or enhanced hybridization signal with the c-*src* gene in the benign tumor DNA sample.

C) ROLE OF AMPLIFICATION AND/OR REARRANGEMENT OF THE *Xsrc* GENE IN THE MELANOMA FORMATION

Protooncogene amplification and/or rearrangement has been reported in a number of tumors and tumor cell lines [7,216]. It was thought that these sorts of phenomena may be playing a role in the melanoma formation in the hybrid fish and could also underlie the observed correlation between *mdl* deregulation and *src* tyrosine kinase activity. To initiate an investigation whether gene amplification and/or rearrangement is playing any role in the melanoma formation in *Xiphophorus* the same two tumor DNA samples as described above (i.e. the first derived from highly malignant tumors and the second from benign tumors) have been assayed by restriction enzyme digestion and hybridization for any readily observable physical changes in the fish c-*src* gene.

In Figure 18a the malignant tumor DNA has been digested with the restriction enzyme Bam HI (along with brain DNA from the *X. maculatus* parent and brain and liver DNA from the hybrid tumor-bearing fish) and hybridized with the *Xsrc*-specific probe B (see Figure 10a). Lane 2 indicates an elevated c-*src* signal compared to the other three lanes. In addition, the band runs slightly faster than the band detected in the other three samples. To control for the relative loading of the DNA digests in the four lanes, the same filter was rehybridized with the *Xhomeobox* specific probe described above (Figure 17b; Appendix 5). In addition, this hybridization serves to control whether the faster migration of the Bam HI hybridizing band detected with the *Xsrc* probe was a gel-running artefact associated with this particular tumor DNA sample. The *Xhomeobox* hybridization illustrates clearly two points (Figure 18b): First, the relative DNA loadings in the four lanes are not equal, with lanes 1 and 2 having equal loadings which are however about half of the loadings in lanes 3 and 4. The same conclusion is reached based upon the

intensity of the hybridization signals in lanes 1, 3 and 4 of part a of Figure 18, excluding a loading artefact as underlying the strong *src* hybridization signal detected. Second, the hybridizing band in all four lanes runs to exactly the same position in the gel, arguing against a gel-running artefact in lane 2 of Figure 18 a.

As with the methylation studies, these analyses were repeated with the benign tumor DNA sample. Figure 18 parts c and d show the results of the *Xsrc* and *Xhomeobox* hybridizations respectively following a *Hind* III digestion of the DNA samples. The *Xsrc* probe detects a single band and the *Xhomeobox* probe detects a double band. First, there is no indication of any rearrangement or other gross change of the *c-src* gene in this particular tumor sample. Second, the lane with the tumor sample contains less DNA than the other three lanes (which was also apparent from the ethidium bromide stained gel -data not shown). When the relative loadings are taken into account it is clear that there is also no indication of any amplification of the *c-src* hybridization signal detected in this DNA sample. Therefore, as with the methylation studies, these results again indicate a general difference between the benign and the malignant tumors concerning specifically the *c-src* gene.

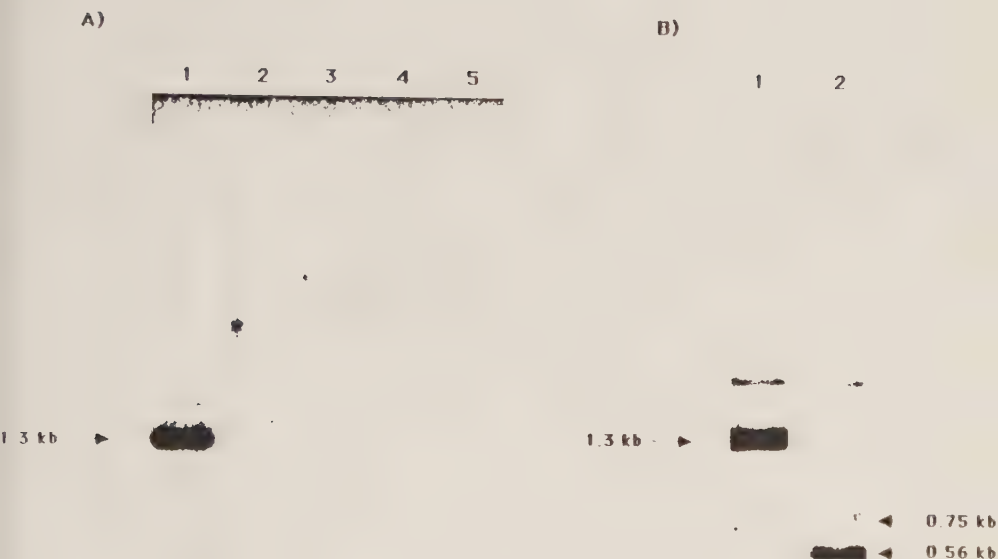


FIGURE 15

Xiphophorus c-src gene does not map to the X-chromosome.

A) *X. maculatus*, *X. helleri* and backcross fish DNA digested with *Sau* 3A and hybridized with the *Xsrc* 0.6 kb *Pst* I/*Xba* I fragment (probe B, Fig. 9a). DNAs: lane 1: *X. maculatus*; lane 2: *X. maculatus* / *X. helleri* / *X. helleri* (multiple backcross to the *X. helleri* parent); lane 3: *X. mac* / *X. hell* / *X. hell* - *X. maculatus* X-chromosome has a translocation involving the *mdl* locus; lane 4: *X. mac* / *X. hell* / *X. hell* - *X. maculatus* X-chromosome has a deletion including the *mdl* locus; lane 5: *X. helleri*. Hybridization and washing under stringent conditions (see Materials and Methods)

B) Repeat of hybridization as described above, but under standard hybridization and washing stringency conditions (see Materials and Methods). Lane 1: *X. maculatus* DNA; lane 2: *X. helleri* DNA.

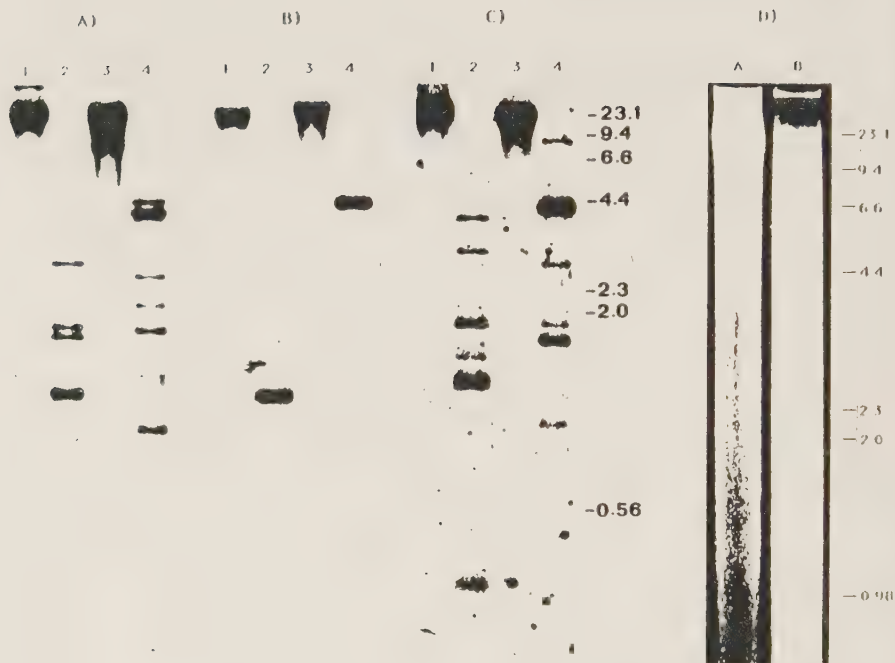


FIGURE 16

Methylation status of *src*-related sequences in *Xiphophorus* testes DNA.

Xiphophorus maculatus testes DNA digested with: lane 1 - Dpn I; lane 2 - Sau 3A; lane 3 - Hpa II; lane 4 - Msp I. Hybridization probes: A) *v-src* 600 bp Pst I kinase domain fragment; B) *Xsrc* Pst I 1.3 kb fragment (probe A, Fig. 9a); C) *v-yes* 1.1 kb Pst I kinase domain fragment. D) Ethidium bromide stained gel following digestion of testes DNA with Hpa II (lane A) and Msp I (lane B). Hybridizations and washings in A) and C) under standard conditions; in B) under stringent conditions (see Materials and Methods).

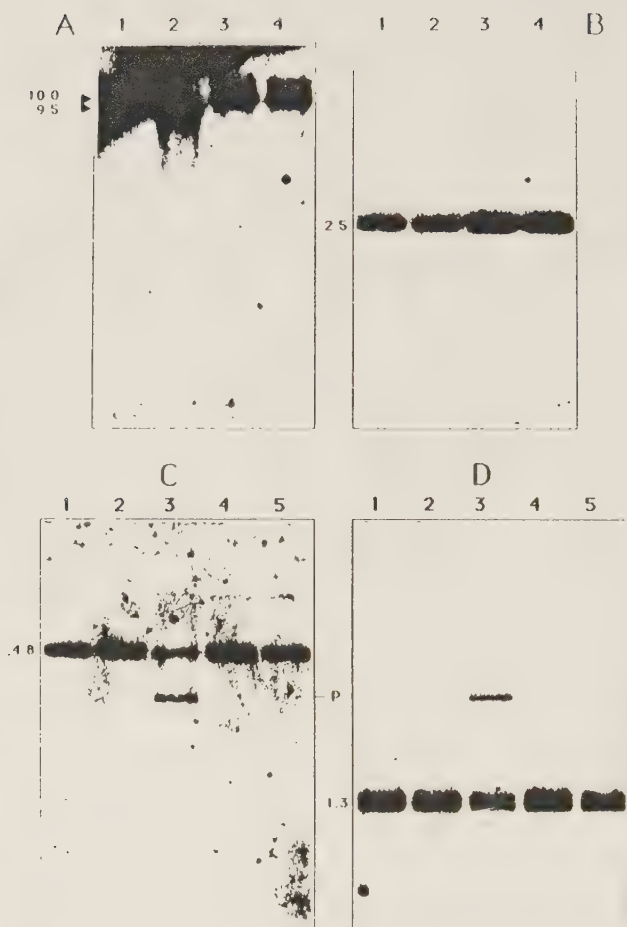


FIGURE 18

Amplification of *c-src* sequences in malignant melanoma DNA.

A) and B): DNA was prepared from: *X. maculatus* brain (lane 1); malignant melanoma (lane 2); brain from malignant melanoma-bearing backcross fish (lane 3), and liver from the same backcross fish (lane 4). Following digestion with Bam HI and transfer to nylon membrane the same filter was hybridized with: A) the *Xsrc* 0.6 kb Pst I/Xba I fragment (probe A, Fig. 9a) and, following removal of bound radioactivity, B) the *Xhomeobox* 1-3 1.85 kb Bam HI fragment.

C) and D): Hybridization of benign melanoma DNA with a *Xsrc* and a *Xhomeobox* probe. DNA was prepared from: *X. maculatus* brain (lane 1); *X. maculatus* liver (lane 2); benign melanoma (lane 3); brain from benign melanoma bearing backcross fish (lane 4), and liver from the same backcross fish (lane 5). Following digestion with Hind III and transfer to nylon membrane the same filter was hybridized with: C) the *Xsrc* 0.6 kb Pst I/Xba I fragment and, following removal of bound radioactivity, D) the *Xhomeobox* 1-3 1.85 kb Bam HI fragment.

P denotes a minor contamination of the tumor DNA sample with a plasmid that is detected by both hybridization probes.

III. RESULTS

3. EVOLUTIONARY ASPECTS OF THE *SRC*/TK GENE FAMILY

A) Divergence rates

Sequence from the kinase domain is now available for the *c-src* and *c-yes* genes of both fish and man [16,303] (i.e. the two orthologous gene pairs fish *src* -human *src* and fish *yes* -human *yes*, and the two paralogous gene pairs fish *src* -fish *yes* and human *src* -human *yes*). The two orthologous gene pairs can be used to determine the genetic distance between these genes in fish and humans. With published sequence data for *c-src* also available from the chicken [305] and the mouse [193], where again one can be sure that one is dealing with strictly orthologous gene pairs, it is possible to calculate the genetic distance measurements for the *c-src* gene compared between these various organisms (Table IV). The values in parentheses are the corrected genetic distance measurements (corrected according to Dayhoff [61]), which takes into account statistically both superimposed and back-mutations that are no longer observable.

TABLE IV

GENETIC DISTANCE MEASUREMENTS ^a (corrected values in brackets)		
	nucleotides (total)	amino acids
FISH SRC - CHICKEN SRC	21.5 (25.0)	9.6 (10.2)
FISH SRC - MOUSE SRC	21.8 (25.5)	9.0 (9.5)
FISH SRC - HUMAN SRC	20.8 (24.2)	8.6 (9.0)
CHICKEN SRC - MOUSE SRC	14.2 (15.6)	2.3
CHICKEN SRC - HUMAN SRC	12.4 (13.5)	1.7
MOUSE SRC - HUMAN SRC	10.4 (11.1)	0.7
FISH YES - HUMAN YES ^b	23.9 (28.5)	6.9 (7.2)

^a = exons 7-12

^b = exons 7, 10 and 11

These data allow the calculation of the approximate gene divergence rates for the *c-src* and *c-yes* genes, knowing from the fossil record that the last common ancestor between humans and bony fish occurred approximately 400 million years (Myr) ago [148]. The number of years of independent evolution between any two extant species is twice the number of years since

their divergence from a common ancestor - i.e. approximately 800 Myr for fish and human. The derived divergence rates (in nucleotide substitutions per 100 residues per 100 Myr; percent accepted mutations per 100 Myr or PAMs/100 Myr) for *c-src* and *c-yes* are 3.0 and 3.6 respectively. At the amino acid level the derived divergence rates (in amino acid substitutions per 100 residues per 100 Myr; PAMs/100 Myr) for the *c-src* and *c-yes* genes are 1.1 and 0.9 respectively. That is, the two genes, following the original gene duplication event, are apparently evolving at a very similar overall rate. These data are summarized in Table V.

TABLE V

sequences compared	DIVERGENCE RATES ^a			
	c-src [*]		c-yes ^{**}	
	nucleo- tides	amino acids	nucleo- tides	amino acids
FISH-CHICKEN	3.1	1.3		
FISH-MOUSE	3.1	1.2		
FISH-HUMAN	3.0	1.1	3.6	0.9
CHICKEN-MOUSE	3.6	0.3		
CHICKEN HUMAN	3.1	0.2		
MOUSE-HUMAN	7.9	0.1		

^a = corrected changes/100 residues/100 Myr

^{*} = exons 7-12

^{**} = exons 7, 10 and 11

Using the *c-src* genetic distance values derived from the organism comparisons in Table IV, it is possible to determine whether the *c-src* gene divergence rate has been constant over evolutionary time. The chicken-human comparison (divergence approximately 215 Myr ago [148]) gives rise to a gene divergence rate very similar to that derived from the fish-human comparison (divergence approximately 400 Myr ago) - 3.1 and 3.0 respectively (Table V). The gene divergence rate calculated from the mouse-human comparison (divergence approximately 70 Myr ago [148]) is a factor of two to three times higher (7.9; Table V). The fish-chicken, fish-mouse and chicken-mouse DNA comparisons are all similar to the fish-human and chicken-human results (see Table V). The amino acid PAM/100 Myr values vary over a ten-fold range (Table V). The comparisons involving the fish protein consistently give the highest divergence rates (1.1-1.3 PAM/100 Myr), whereas the mouse-human comparison gives the lowest divergence rate (0.1 PAM/100 Myr). The chicken-mouse and chicken-human comparisons give intermediate values (0.2 and 0.3 PAM/100 Myr).

B) *src/yes* gene duplication

By comparing the sequences of the two paralogous gene pairs (i.e. human *c-src* and *c-yes* ; fish *c-src* and *c-yes*), it is possible to make an approximation of genetic distance between the two genes in the respective organism. This can be done by again correcting the absolute differences observed between the genes by using the mutation probability matrix of Dayhoff. The results of these comparisons are shown in Table VI.

TABLE VI

GENETIC DISTANCE MEASUREMENTS (corrected values in brackets)

	<u>nucleotides</u> <u>(total)</u>	<u>nucleotides</u> <u>(3rd position)</u>	<u>amino acids</u>
HUMAN SRC - HUMAN YES ^a	27.6 (34.2)	65.4 (154.0)	10.0 (10.7)
FISH SRC - FISH YES ^b	26.0 (32.0)	59.8 (118.0)	13.0 (14.2)

^a = exons 7-12

^b = exons 7, 10 and 11

These calculations can be made over a range of genetic distance and gene divergence parameters (eg. total nucleic acid substitutions, substitutions at the first, second or third position of the codon, amino acid substitution, etc). Coupling this with the calculation of the divergence rates for the two genes, it is possible to extrapolate back and determine the approximate time of the *src/yes* gene duplication from an ancestral sequence. An example of one of the extrapolations is presented in Figure 19 for both the fish and human *c-src* and *c-yes* gene pairs based upon total nucleic acid base substitutions. These results suggest a *src/yes* gene duplication event occurring somewhat over 500 Myr ago. When similar extrapolations are done with both the human and the fish data for several different genetic distance parameters (see Table VI), the average time arrived at is approximately 550 Myr ago (Table VII). Figure 20 presents this result against both an absolute time scale as well as including for reference commonly accepted divergence times for several major evolutionary lineages. **TABLE VII**

CALCULATED TIMES FOR THE SRC/YES DUPLICATION (Myr)

	<u>total</u> <u>nucleotides</u>	<u>nucleotides</u> <u>(3rd position)</u>	<u>amino acids</u>
HUMAN SRC - HUMAN YES	515	625	535
FISH SRC - FISH YES	495	480	650

AVERAGE: 550 ± 50

These calculations of genetic distance between the *c-src* and *c-yes* genes in humans and fish demonstrated that the fish *c-src* and *c-yes* genes are further diverged at the amino acid level (exons 7, 10 and 11) than are the two human genes over the same three exons (12.2% versus 5.5% amino acid substitutions respectively). However, most of these changes in the fish genes are of a conservative nature so that when one considers amino acid identity plus conservative substitutions the genetic distance measurements are somewhat closer to one another (5% and 3.9% respectively). The higher genetic distance measured between the fish *c-src* and *c-yes* genes is predominantly due to divergence in exon 7. Within exon 7, the *Xyes* gene is closely related at the amino acid level to both the human SRC and YES genes (88.5% and 92.3% respectively), whereas the *Xsrc* gene is reduced in relationship to both the human SRC and YES genes (88.5% and 88.5% respectively), and the *Xyes* gene (80.8%). The human SRC and YES genes are 96.1% identical at the amino acid level in exon 7. These results suggest that the sequence divergence has occurred predominantly in the *Xsrc* gene during the independent evolution of that gene in the teleost lineage.

C) Evolution Southern

As outlined in Results Section I, Southern blots of *Xiphophorus* genomic DNA were able to differentiate between fish genomic sequences more related to the *v-src* probe and those more related to the *v-yes* probe. Because the preceding calculations regarding the *src/yes* gene duplication suggested a correlation between this duplication and the emergence of the first protochordates, an attempt was made to test for the presence of separate *src* - and *yes* -related sequences by Southern blots from extant species thought to approximate primitive proto-chordates.

In Figure 21 are shown the results of an experiment designed to test the general feasibility of this experimental plan and includes a survey of organisms believed relevant to the evolution of the vertebrates. Most evolutionary histories for the phylum Chordata assume that *Amphioxus* (Subphylum Cephalochordata) arose from the larval form of an organism similar to the present-day tunicates. The jawless fishes (present-day hagfishes and lampreys; Subphylum Vertebrata, Class Agnatha) are believed to have evolved from an ancestral form that resembled the present-day *Amphioxus* [224]. It was reasoned that in such hybridizations a sequence more related to the *v-yes* probe than the *v-src* probe would result in either a stronger relative signal for a particular band compared to the *v-src* hybridization, or, alternatively, that a new band would be detected with the one probe not detected by the other. A band that hybridizes equally strongly with both probes is indicative, under the stringency conditions for hybridization and washing selected, of a genomic sequence related approximately equally to both probes. The analysis of DNAs derived from organisms belonging to lineages that diverged before the proposed time of the *src/yes* duplication should serve as a test for the accuracy of this Southern blot assay for separate *c-src* and *c-yes* sequences within a genome - i.e. in these cases separate *c-src* and *c-yes* sequences should not be detected.

In the human, chicken, and fish DNA samples it is possible to detect bands that hybridize with only one probe or predominantly with only one and not (or reduced) with the

other, indicating the presence within these genomes of separate sequences more closely related to the v-src probe and others more closely related to the v-yes probe. The *Amphioxus* DNA sample gives a complex array of hybridizing bands with the v-src probe, one of which also hybridizes (apparently slightly stronger) to the v-yes probe. The tunicate (*Ciona intestinalis*) DNA sample gives rise to separate v-src and v-yes hybridizing bands following digestion with several different restriction enzymes (marked by arrows in Figure 21), consistent with the presence of separate sequences more src-related and yes-related within this genome. The exceptions to the above observations are the lamprey (*Lampræta planeri*) and the shark DNA samples. The lamprey DNA gives a complex but relatively normal pattern of bands on the autoradiogram following hybridization with the v-src probe, but to date has given no readily detectable signal with the v-yes probe. This particular hybridization has been repeated several times under various conditions, with the results always being negative with the v-yes probe. The shark DNA sample gives rise to a complex pattern of strongly hybridizing bands following hybridization with the v-yes kinase domain probe, but in comparison to the other samples gives rise to only one or possibly two weakly hybridizing bands with the v-src hybridization probe - both of which are also detected with the v-yes probe.

In Figure 22 are presented the results of hybridizations with the two viral probes onto DNA derived from the fruit-fly *Drosophila melanogaster* (Phylum Arthropoda, Class Insecta), a crab (*Cancer pagurus*; Phylum Arthropoda, Class Crustacea), and, as representative for a more primitive evolutionary level, the tapeworm *Diphylobothrium dendriticum* (Phylum Platyhelminthes, Class Cestoda). In the crab and tapeworm hybridizations a single hybridizing band is detected with the v-src probe, and no signal is detectable with the v-yes probe. The *Drosophila* hybridization with the v-src probe detects several src-related sequences whereas the v-yes probe detects one band which is also detected with the v-src probe.

D) src-based organismal phylogeny

The sequence data from the fish c-src gene can be combined with the sequence data available for the human, mouse and chicken c-src genes to prepare a phylogenetic tree based upon c-src sequences. In addition, a small amount of sequence is available for the *Xenopus* c-src gene (only about 40 amino acids from exons 11 and 12 [295]), which has been excluded because this very small sample set means that the *Xenopus* placement in the tree would be subject to considerable uncertainty. From these data a distance matrix can be prepared and a phylogenetic tree derived based upon the principle of least parsimony (see Materials and Methods). Figure 23 presents a distance matrix based upon the amino acid sequence data for the human, mouse, chicken and fish c-src genes, and the least parsimonious tree. It is generally found that such trees accurately depict the branching order as predicted based upon more classical paleontological analyses, and this is found to be the case for the tree presented in Figure 23. When one includes the sequence data available for the two *Drosophila* c-src genes [118,283] and from a sponge src-related gene (S. Otilie, unpublished; see below), one maintains the general topology of the tree as presented, and the *Drosophila* and sponge sequences enter the tree basically as outgroups, but

in an order as expected based upon sequence available and paleontological considerations (i.e. the sponge sequence most distant from the top of the tree - see Figure 3 in Appendix).

Such sorts of phylogenetic analyses are best not based upon sequence derived from only one gene, and the derived trees become robust as more independent sequence data sets are incorporated, and as sequence from more organisms is included. Therefore, as part of a longer term project looking further at evolution of the *src* gene family, *src* -related sequences have been isolated from a genomic library of *Xenopus laevis* (provided by Prof. W. Knöchel, Berlin) and from a genomic library of the nematode *Caenorhabditis elegans* (provided by J. Sulston, MRC, Cambridge). These clones are still at an initial stage of characterization, and Figure 24 shows the results of a 'giant plaque' and restriction/hybridization analysis of several of the *Xenopus* clones. These results demonstrate two things: first, as observed for *Xiphophorus*, the frog genome clearly contains sequences which differ in their relative sequence similarity to several members of the *src* gene family (eg. clone 1-1, Fig. 24a); and second, the strongly *src* -related clones fall into two major groups (see for example the Hind III digests of clones 5-3 and 7-1, Fig. 24b) based upon these preliminary restriction enzyme analyses.

As a first approach towards expanding the sequence data set available from *Xiphophorus* for genes other than those deriving from the *c-src* gene family, the *c-abl* gene of *Xiphophorus* was initially chosen. The *c-abl* gene was of interest for two reasons: 1) the gene also encodes a tyrosine kinase, but belongs to a different gene family than *c-src* [143], and an evolutionary analysis may uncover additional relationships between the two gene families, and 2) the *c-abl* gene is also a gene highly conserved during evolution, with *c-abl* genes having been isolated from (among others) *Drosophila* [132] and the nematode *C. elegans* [98], which means that a sequence data set for the gene is accumulating. Several *abl* -related sequences have been isolated from the *X. maculatus* gene library, and these have been subjected to a similar preliminary analysis as with the *Xenopus src* -related clones. In Figure 25a are shown the results of a 'giant plaque' analysis of five of these clones, demonstrating that one clone (ABL-5) is most closely related to the *v-abl* probe (the other clones are *abl* cross-reactive under less stringent washing conditions). Figure 25b shows the results of a preliminary restriction enzyme analysis of ABL-5, whereby an approximately 1 kb Hae III fragment from ABL-5 contains the sequence responsible for the *v-abl* hybridization.

E) Primordial *c-src* sequences

The data presented so far concerning the evolution of the *c-src* gene and the *src* gene family raises the question of the origin of this particular gene and gene family. Several observations suggested that the *c-src* gene is of extremely ancient origin. Scharf and Barnekow could detect a tyrosine-specific protein kinase activity in the freshwater sponge *Spongilla lacustris* using a rabbit antiserum raised against the viral *src* protein [22,262]. In addition, Southern and Northern analyses succeeded in detecting *src* -related sequences in both the genomic DNA and the RNA of the sponge [22], suggesting that a *src* -related gene was present, was expressed and coded for a tyrosine kinase. As the phylum Porifera is of ancient origin (estimates average approximately 10^9 years) [262] and represents the most simple metazoan animals

[245,310], it was thought that the *src*-related gene in the sponge genome might represent a primordial *c-src* sequence.

A genomic library of *Spongilla lacustris* DNA was prepared in the lambda vector EMBL 4, and screened with the *v-src* kinase domain probe essentially as described for the *Xiphophorus* library, except that the filters were washed at reduced stringency (see Materials and Methods). Three positive clones remained after initial picking, rescreening and plaque purification. These three clones have been subjected to restriction enzyme mapping and hybridization with the *v-src* kinase domain probe, and the results are presented in Figure 26a. The three clones appear to be of independent origin with no obvious region of overlap yet detected between any of them. Clone 4-1 is clearly the most strongly hybridizing of the three, with clones 2-2 and 1-1 giving approximately equal, albeit reduced signals. Figure 26b presents an example of a hybridization of clone 4-1 with the *v-src* kinase domain probe, demonstrating the specificity of the signal detected. The further analysis of these clones and in particular clone 4-1, including sequencing, is being continued in our laboratory by Sabine Otilie as part of her ongoing Ph.D thesis.

Schartl and Barnekow, as part of a wide-ranging survey of *src*-related tyrosine kinase activity in the animal kingdom, also detected a relatively high activity in the coelenterate *Hydra attenuata* [266]. Figure 27 presents the results of hybridizing two duplicate *Hydra viridissima* genomic Southern blots with two different *src* probes - the *v-src* kinase domain probe (Fig. 27a and b) and a fragment derived from the kinase domain of the *Xiphophorus c-src* gene (Fig. 27c). *Hydra viridissima* was chosen over *H. attenuata* for this analysis because the *H. viridissima* genome is approximately three times smaller than the genome of *H. attenuata*, which is at least as large as the mammalian genome. Both probes hybridize with more than one band, and the two predominant bands detected by the *Xiphophorus* probe are a subgroup of the bands detected with the *v-src* probe. Experiments aimed at the cloning of *src*-related sequences from the *Hydra viridissima* genome, utilizing sub-genomic libraries prepared from the hybridizing bands detected in Figure 27, are presently in progress.

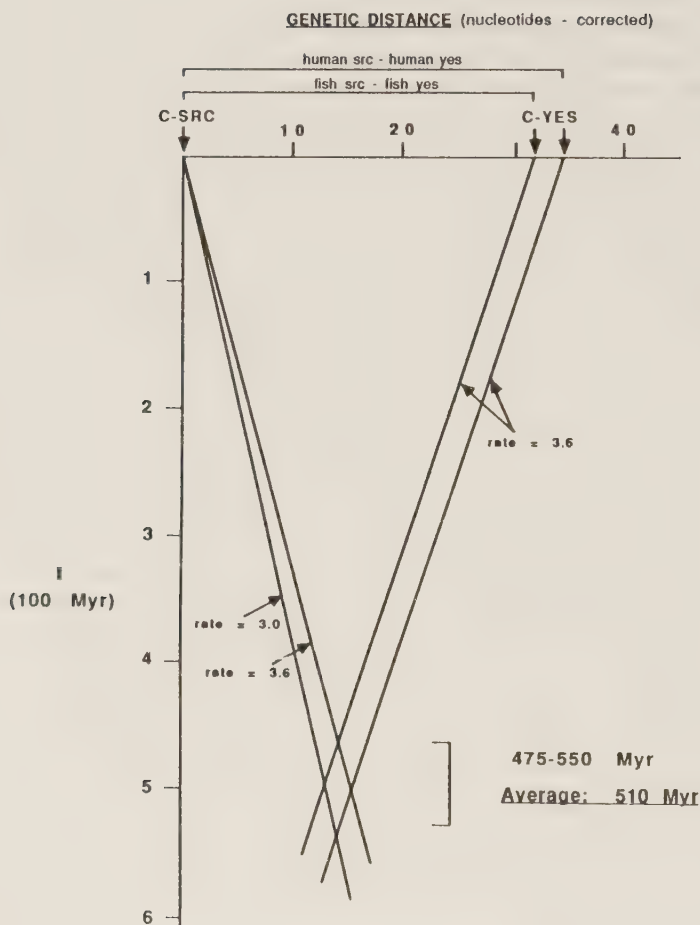


FIGURE 19

Calculation of the time of the *src/yes* gene duplication based upon total nucleotide genetic distance values. For the *c-src* gene the range of gene divergence rates calculated (see Table V; excluding the rate based upon the mouse/human comparison) are plotted against the *c-yes* gene divergence rate based upon the fish/human comparison (Table V). Extrapolations based upon genetic distance measurements for both the fish and human *src-yes* comparisons are shown.

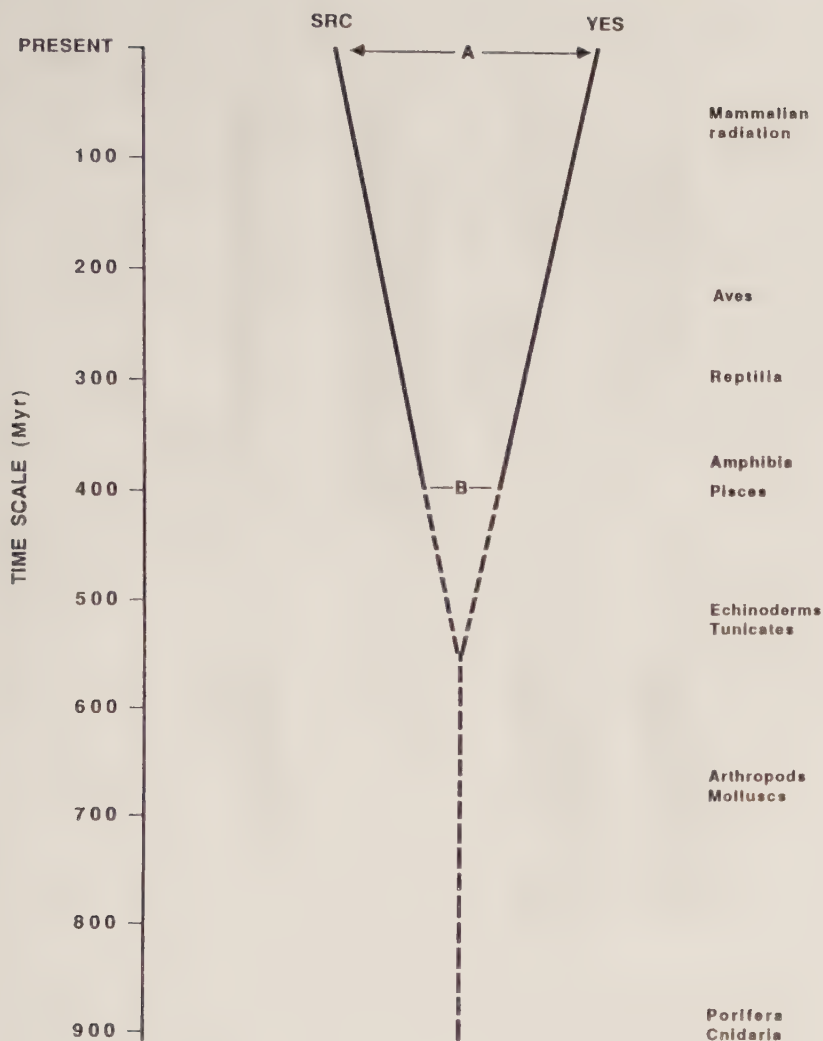


FIGURE 20

Schematic presentation of the proposed *src/yes* gene duplication against an absolute time scale (left) and including the approximate time of divergence of major lineages (right). A = the corrected distance between the present-day fish and human *c-src* and *c-yes* genes. B = the presumed genetic distance between the *c-src* and *c-yes* genes at the time of the divergence of the fishes. Solid line: based upon available sequence data. Dashed line: hypothetical.

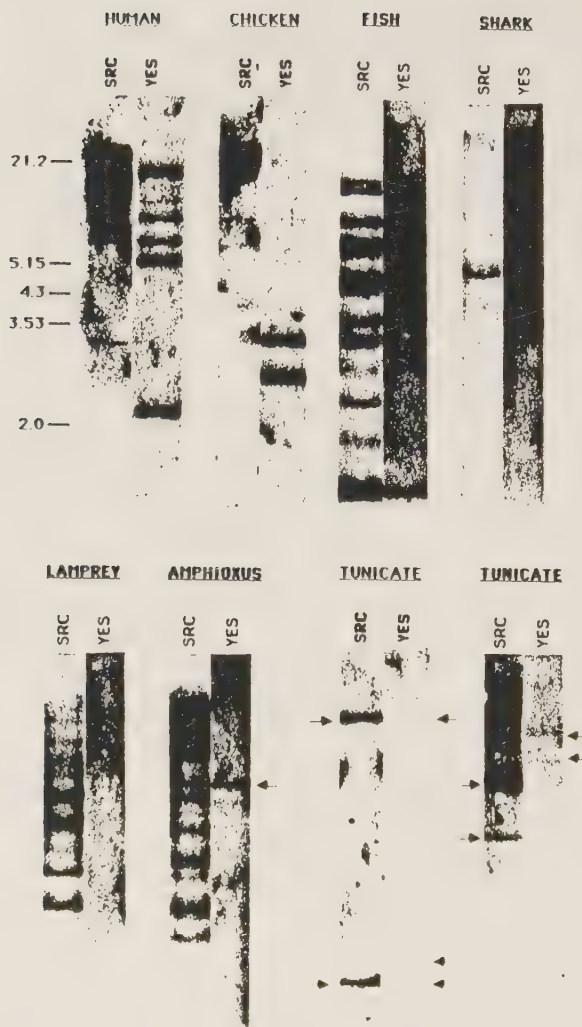


FIGURE 21

Evolution Southern blot hybridizations.

DNA from various organisms pertinent to the evolution of the vertebrates were digested with Pst I, transferred to a nylon membrane, and hybridized successively with the v-src 600 bp Pst I fragment (kinase domain) and the v-yes 1.1 kb Pst I fragment (kinase domain). Tunicate: left panel - Pst I; right panel - Eco RI. Standard stringency conditions (Materials and Methods).

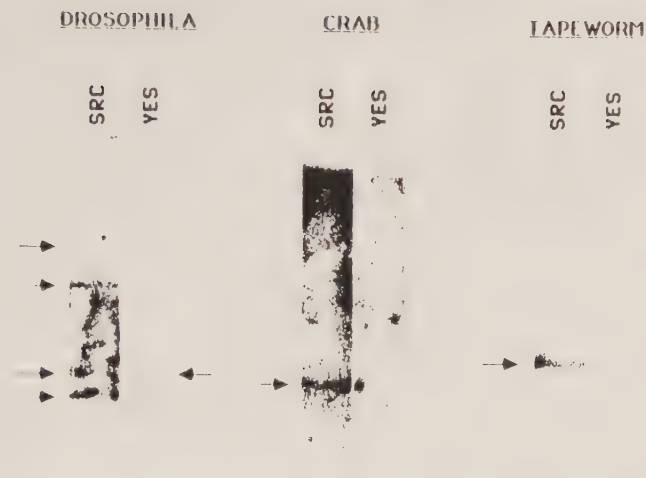


FIGURE 22

Evolution Southern blot hybridizations

DNA from various organisms whose lineages diverged prior to 500 Myr ago from that lineage leading eventually to the modern-day vertebrates. DNAs were digested with Pst I, transferred to a nylon membrane, and hybridized successively with the v-src and v-yes kinase domain probes.

Standard hybridization stringency; washings were performed at 50°C in 1X SSC, 1% SDS.

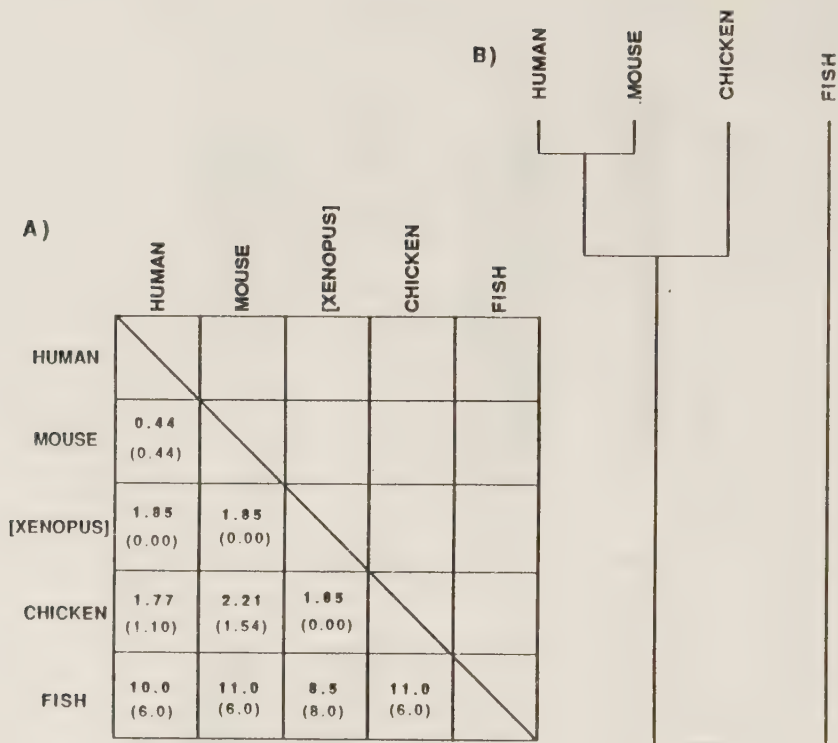


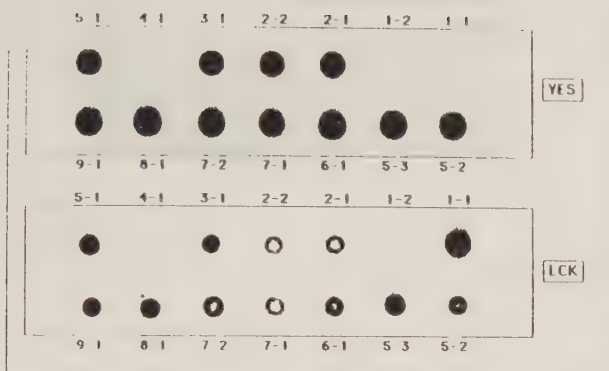
FIGURE 23

src organismal phylogeny.

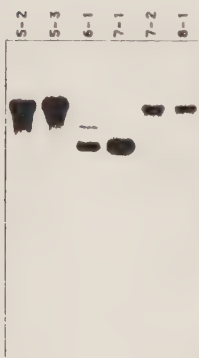
A) Distance matrix (amino acid residues). Marked are the corrected PAM scores for amino acid residue identity (bold type) and for identity plus conservative changes (in brackets). PAM = percent accepted point mutation - i.e. in this case simply percent divergence at the amino acid level between the various sequences. These values have been statistically corrected for superimposed or back-mutations which have occurred but are no longer able to be scored [see Dayhoff, 1978]. The *Xenopus* values are very tentative as they are based upon only a small amount of available sequence data

B) The least parsimonious phylogenetic tree based upon the distance matrix in A). A constant evolutionary rate was assumed for the *c-src* gene, which does not affect the general topology of the tree. *Xenopus* has been excluded from the tree as the small size of the data set does not allow a rigorous enough placement within the tree.

A)



B)



C)

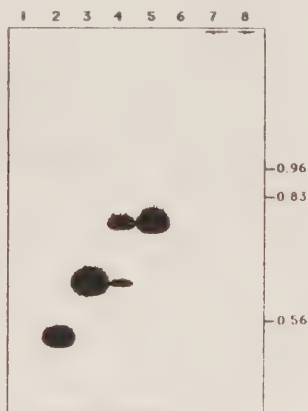


FIGURE 24

Xenopus src-related clones.

A) Giant-plaque analysis of 14 *src*-related *Xenopus* clones. Hybridizations were performed with the *v-yes* and *c-lck* kinase domain probes under standard stringency conditions. Note the very low reactivity of clone 1-1 with the *v-yes* probe but strong reactivity with the *c-lck* probe. Conversely, clones 2-1, 2-2 and 7-1 are weakly reactive with the *c-lck* probe but give strong signals with the *v-yes* probe.

B) Six *src*-related *Xenopus* clones digested with *Hind* III, transferred to a nylon membrane, and hybridized with the *v-src* 600 bp *Pst* I fragment (standard stringency conditions). Note that the clones appear to fall into 2 classes. Absolute strength of hybridization signal not directly comparable due to unequal DNA loadings in the gel.

C) Fractionation of the *src*-reactive fragments from clone 5-3. Lambda DNA digested with *Hae* III, run on a 0.6% low melting point (LMP) agarose gel, and the gel size-fractionated by slicing into bands. Gel pieces were then melted, and an aliquot from each band loaded onto a standard agarose gel. Following transfer to a nylon membrane, hybridization was carried out with the *v-src* 600 bp *Pst* I fragment under standard conditions. Those gel fractions testing positive were remelted and sub-cloning in the presence of LMP-agarose performed (see Materials and Methods).

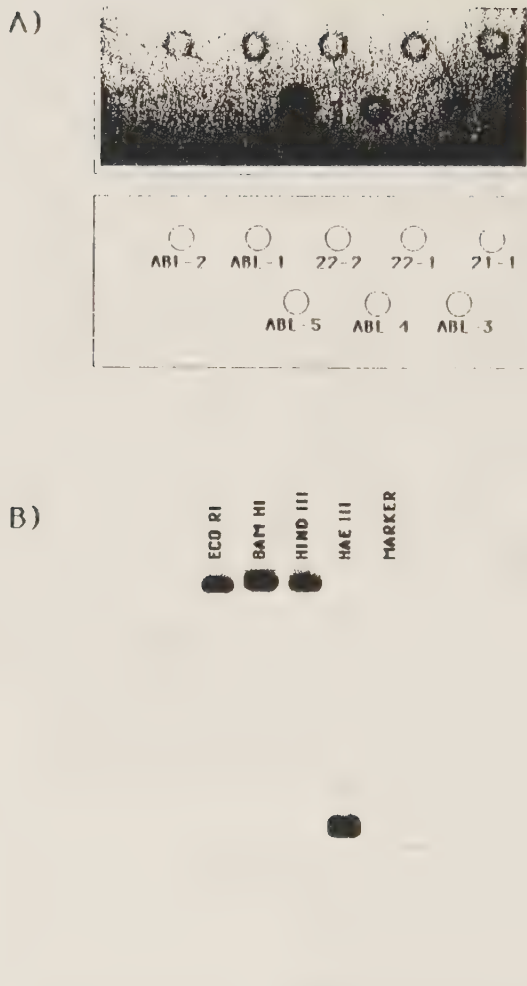


FIGURE 25

Xiphophorus abli-related clones.

A) Giant-plaque analysis of the 5 *abli*-related clones. Hybridization performed with a *v-abl* kinase domain probe. Standard hybridization stringency conditions; washing carried out at 65°C in 1X SSC, 1% SDS. Clones 21-1, 22-1 and 22-2 are *Xiphophorus src*-related clones. Note the strong reactivity of clone ABL-5 under these increased stringency conditions.

B) ABL-5 lambda DNA digested with various enzymes, transferred to a nylon membrane and hybridized with the *v-abl* kinase domain probe under standard stringency conditions (Materials and Methods). Note: 1) the bulk of the hybridization signal can be localized to a Hae III fragment of approximately 1 kb, and, 2) the wild-type lambda DNA/Hind III marker lane is free from any cross-hybridization signal.

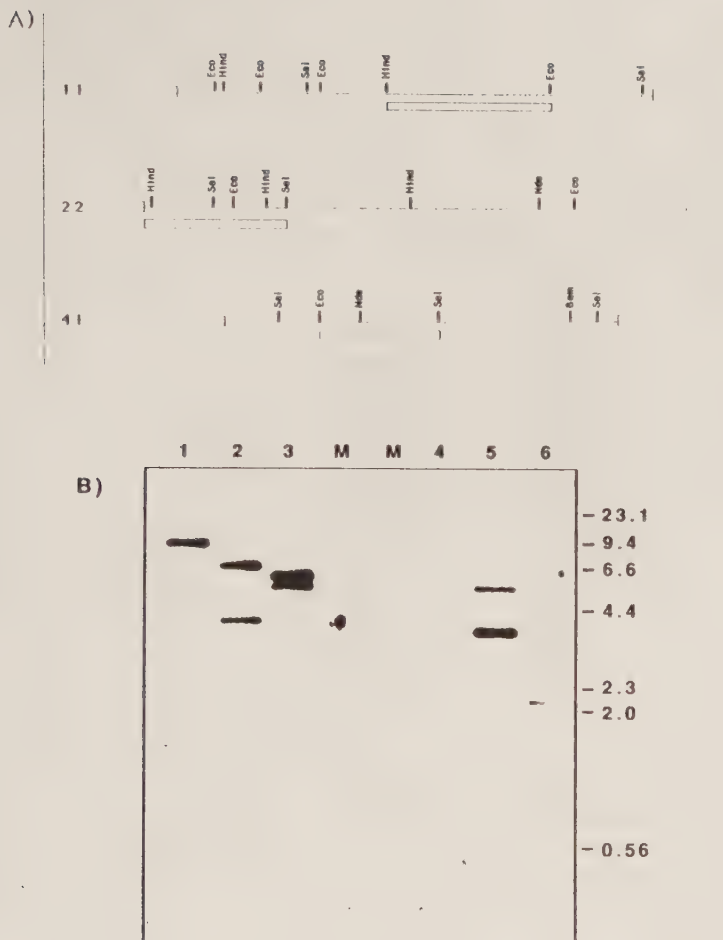


FIGURE 26

Sponge src-related clones.

A) Preliminary restriction maps of the three *Spongilla lacustris* src-related clones.

B) Hybridization of sponge-src clone 4-1 with the v-src 600 Pst I fragment (under standard stringency conditions) after digestion of the recombinant lambda DNA with the following enzymes used in restriction mapping: 1) Bam HI + Eco RI; 2) Bam HI + Sal I; 3) Nru I + Sal I; 4) Nar I + Nde I; 5) Nde I + Sal I; 6) Nru I + Nar I. M denotes lambda wild-type digestions run as molecular weight markers.



FIGURE 27

Hydra viridissima genomic Southern blot hybridized with *src* probes.

Hydra viridissima (lane 1) and *Xiphophorus helleri* (lane 2) DNAs were digested with Pst I and two duplicate pairs of samples transferred to a nylon membrane.

A) and B) One of the duplicate filters was hybridized with the *v-src* 600 bp Pst I fragment under standard hybridization stringency conditions and washed at: A) 50°C in 1X SSC, 1% SDS, and B) 60°C in 1X SSC, 1% SDS.

C) The other duplicate filter was hybridized with the *Xsrc* 600 bp Pst I/Xba I fragment entailing exons 8, 9 and 10 (probe B, Figure 10a) under standard stringency conditions. Predominant hybridizing bands in the *Hydra* DNA sample are marked with arrows and their sizes (in kb) given.

Note the 3.5 kb band which gives a strong hybridization signal with both *src* probes

IV. DISCUSSION

1. FISH SRC/TK FAMILY

The sequence of the kinase domain of the *Xiphophorus maculatus* *c-src* gene is presented. The conservation of all amino acid residues diagnostic of protein kinases in general, as well as the two motifs diagnostic for tyrosine-specific protein kinases leaves little doubt that this gene does indeed encode a tyrosine kinase. This represents the first report of sequences from the *src* gene-family of a fish, and in addition are the first sequences from a fish gene encoding a protein tyrosine-kinase.

A number of hybridization analyses support the proposal that additional *src* -related sequences exist within the *Xiphophorus* genome (eg. see Figures 2, 5, 6 and 9). For example, the clone 22-1 kinase domain has been partially sequenced (G. Hannig, unpublished) and has been shown to be derived from the *Xiphophorus* *c-yes* gene (Table III). Likewise, preliminary analyses of clone 22-2 are consistent with this clone deriving from the *Xiphophorus* *c-lck* gene. The question whether the *Xiphophorus* genome contains the full complement of *src* -family members as described to date in higher vertebrates (eg. chicken, mouse and man) must await the in-depth analysis of the other classes of *src* -related clones isolated from the *X. maculatus* genomic library. It is quite clear however that a number of the bands detected in a genomic Southern of fish DNA with a probe from the *v-src* kinase domain derive from the genomic sequences included in clones 19-4 (*Xsrc*), 22-1 (*Xyes*) and 20-2 [Table II]. It is to be expected that the other bands detected in the genomic Southern represent yet further members of the fish *src* gene-family, some (or all) of which may be included in the yet to be fully characterized *src* -related clones. This complex banding pattern in *Xiphophorus* following *v-src* hybridization under moderate stringency conditions has been interpreted by Anders and co-workers as indicating multiple copies (and different alleles) of the *c-src* gene [12,13], a conclusion that ignores the fact that *c-src* is a unique gene in all other vertebrates looked at and is but one member of a closely related family of genes. The results reported here demonstrate the incorrectness of this proposal.

The genomic Southern hybridizations of *X. maculatus* and *X. helleri* DNA with the *v-src* kinase domain probe detected some differences in the banding pattern obtained, although a number of the bands were common between the two DNA samples (Figure 2a). Hybridizations of these DNAs digested with a number of different restriction enzymes with fragments derived from the *c-src* genomic locus of *X. maculatus* resulted in basically the same hybridizing band (or bands) being detected in both *X. maculatus* and *X. helleri* (Figure 7b and data not shown). This suggests that the *X. maculatus* and *X. helleri* *c-src* genes are very closely related, so much so in fact that the sizes of the homologous introns in the respective genes are of the same, or very nearly the same length. This does not mean however that there is necessarily a high degree of sequence conservation between the two genes, particularly in the introns. Experiments aimed at distinguishing the *X. maculatus* *c-src* gene from that of *X. helleri* by way of restriction

fragment length polymorphisms (RFLPs) demonstrated that under very high stringency conditions it was possible to differentiate between the two genes with a *X. maculatus* c-src probe without recourse to a RFLP (Figure 15). This suggests, nevertheless, that the *X. maculatus*/*X. helleri* divergence was relatively recent, since intron sequences are known to evolve quite rapidly, both in terms of sequence and size.

The expression of the *Xsrc* gene has been the subject of the Ph.D thesis of W. Maeuoler and the ongoing thesis of F. Raulf of our laboratory. In summary, the level of mRNA detected in various adult organs, at different embryonic stages and in the melanoma [197,246] parallels the level of pp60^{c-src} tyrosine-specific protein kinase activity as determined earlier for *Xiphophorus* [23,262,264]. These data also support the contention that clone 19-4 does indeed encode a tyrosine kinase, and that clone 19-4 is in fact the *Xiphophorus* c-src gene, as is suggested by the level of nucleic acid and amino acid sequence similarity [Table III]. Final proof that clone 19-4 represents the c-src gene would come from the use of a 19-4 exon-specific probe to select mRNA which following translation *in vitro* gives rise to a tyrosine kinase activity reactive with the rabbit antisera used in the original kinase assays of Scharff and Barnekow.

A cDNA clone of the c-src gene has been isolated from a library prepared (Clontech) in lambda gt10 from poly A⁺ mRNA of a *Xiphophorus* embryonic cell line screened with a fragment from clone 19-4 (Probe C, Figure 10a; F. Raulf, unpublished; Robertson, S.M. et al, submitted). Most of this clone has been sequenced (F.R., unpublished; Robertson, S.M. et al, submitted), and the sequence available from the region of exons 7-11 confirms by and large the sequence derived from the genomic clone. In addition, the exon 12 sequence for *Xsrc* (see Fig.'s 10b & 14) was derived from the cDNA clone. The cDNA sequence that coincides with the genomic sequence confirms the placement of the introns in clone 19-4, as inferred from sequence comparison to the chicken and human genes (Figure 13).

This cDNA clone can now be used to address one puzzling result from the hybridization analyses of the src-related clones- namely, the apparent lack (see Figure 4) or low level of hybridization (see Figure 8b) of these clones, including the c-src clones, to a probe from the extreme 5'-end of the v-src gene. This could be the result of: 1) the high degree of sequence divergence in this region of the gene which would give rise to a probe that would be quite specific for the c-src gene [338], whereby the hybridization and washing conditions utilized in Figure 3 may have been too stringent to detect the fish c-src gene; 2) the 5' end of the src-related genes may not be contained within the available genomic clones, or 3) that during the growth of these particular clones the 5' end of the genes have been lost by recombination. Bishop and co-workers have reported exactly this phenomenon when trying to grow up lambda clones of the chicken c-src gene [235], and have interpreted this as being due to the presence of repetitive sequences in the 5' region of the gene and upstream of the gene that facilitate recombination and removal of the intervening sequences. The placement and orientation of the v-src related sequences within clone 19-4 (*Xsrc* - see Figure 9) indicate that approximately 7.8 kb are available to house the fish sequences homologous to the 350 and 250 bp v-src probes (see Figure 1, Appendix). The hybridizations with the v-src 350 bp probe suggest that the homologous region within the fish

genome is relatively small (Figure 8A-d, lanes 1 and 5). Unless the fish *c-src* gene has very large introns between the first two or three exons then one would expect the 5' end of the genomic locus to be present within clone 19-4. However, the chicken *c-src* gene does indeed have larger introns in the 5' half of the gene compared to the 3' half (exons 2-6: 4193 bp exons plus introns; exons 7-12: 1326 bp exons plus introns) [305], and the situation may be analogous in the fish *c-src* gene.

Option 1) above receives support from the fact that faint but distinct hybridization signals have on occasion been detected when hybridizing digests of clone 19-4 to the *v-src* 5'-probe (Figure 8f). It is interesting in this context to note that Steele, in presenting a preliminary characterization of the *Xenopus laevis* *c-src* gene, also reported that this gene was non-reactive with the extreme 5' *v-src* probe [295]. This problem can be more directly approached by: 1) hybridizing the cDNA clone (which is known to contain the fish sequences homologous to the *v-src* 5'-probe) with the *v-src* probe under various hybridization and washing stringency conditions, and 2) hybridizing the genomic clone 19-4 with probes derived from the 5' end of the cDNA clone, which will clearly demonstrate whether clone 19-4 contains the corresponding sequences. Hybridization of the cDNA clone with the *v-src* 400 bp fragment (i.e. *v-src* 5' probe - see Fig. 1, Appendix) under stringency conditions identical to those used in Figure 8f gives rise to weak hybridization signals with similar intensities to those detected in Figure 8f (data not shown). These results support option 1 outlined above, and are consistent with clone 19-4 containing the entire *Xiphophorus c-src* gene.

If indeed the full *c-src* gene plus 5' sequences are present in this genomic clone (19-4), then a probe from the 5'-end of the cDNA clone should allow primer extension analyses on total cellular RNA, with the resultant mapping of the transcription start site within the genomic clone and characterization of 5' upstream regulatory sequences. These sequences will be particularly significant for the analysis of the regulation of *c-src* expression in relation to the melanoma formation in the hybrid fish (see below).

The sequence of the *Xsrc* gene suggests that the regulation of the activity of the enzymatic activity of the pp60^{*c-src*} protein in the fish cells may well be analogous to that found for the chicken pp60^{*c-src*} protein. That is, the fact that the tyrosine residue near the C-terminus (Tyr 527 in chicken pp60^{*c-src*}), shown to play a major role in the regulation of tyrosine kinase activity in the chicken *c-src* protein [55,142,241], is conserved in the human *c-src* and *c-yes* proteins as well as the fish *c-src* protein argues strongly that all these proteins are regulated through this residue in a similar fashion. Indeed, this tyrosine residue is found in an analogous position in all *src*-family members in all organisms tested so far, even the *Drosophila c-src* gene *Dsrc* 64B [144,312] - but not *Dsrc* 28C [118]. In addition, the sequence of exon 2 derived from the cDNA clone shows that the pp60^{*c-src*} protein of the fish contains a glycine residue after the initiator methionine (F. Raulf and S.M.R., unpublished) which, by analogy to the situation for the chicken pp60^{*c-src*} protein, is probably myristylated following the removal of the initiator methionine [57]. All known members of the *src* gene family (with the exception of the chicken *c-trl* gene where it has not yet been determined) contain a glycine residue following the initiator methionine. Therefore, the fish *c-src* protein is potentially also

localized at the under-surface of the plasma membrane, which has been shown for the chicken protein to require the post-translational myristylation. In fact, the chicken viral protein is non-transforming unless myristylated on this glycine residue and localized under the plasma membrane [57].

It has been proposed by Bernardi and colleagues that a distinction exists between genomes of warm-blooded and cold-blooded animals in that the warm-blooded animals have a tendency towards a higher bias for G or C as the third or 'wobble' base of the codon [25,26,206]. This correlates with an increased thermal stability of the proteins [26]. The tabulation of the codon usage for the *Xiphophorus* *c-src* gene (cold-blooded) in comparison to the *c-src* genes of chicken, mouse and man (all warm-blooded) shows results consistent with this proposal (see Appendix 2). This effect is predominantly seen in the codons for aspartic acid (ASP), valine (VAL), arginine (ARG) and serine (SER). The underlying reason for this observation is unknown, although it conceivably may be due to the stronger G-C bond conferring a greater thermal stability to the DNA molecule, which may have a selective advantage in the warmer intracellular environment of the homiotherms.

IV. DISCUSSION

2. INVESTIGATIONS OF THE ROLE OF THE C-SRC GENE IN THE CROSSING CONDITIONED MELANOMAS OF *XIPHOPHORUS*

Schartl *et al* reported in 1982 the correlation between the degree of deregulation of the *mdl* locus and the level of expression of the *c-src* gene, as assayed by tyrosine-specific protein kinase activity reactive with an antiserum raised in rabbits infected with Rous sarcoma virus [263]. Specifically, these workers found a basal level of *src* tyrosine kinase activity in all *Xiphophorus* strains tested, including those normally used in the production of crossing-conditioned melanomas. In the F_1 hybrid fish with the mild melanosis ('benign melanoma' according to Anders and co-workers) there is an increase in the level of the tyrosine kinase activity, and this increase is specifically detected in both the tumor tissue and the brain of these fish. In the backcross generation, the 50% of the fish with no spots have the basal level of kinase activity in brain tissue, the 25% with the mild melanosis (i.e. phenotypically identical to the F_1 fish) have in both brain and tumor tissue the elevated kinase level detected in the F_1 fish, whereas the 25% with malignant melanoma have a further elevation of the kinase activity over that found in the F_1 fish, again in both brain and tumor tissue. The reason for this co-regulation between *c-src* activity in both the tumor tissue and brain is unknown, but was used by Schartl and co-workers as a convenient indirect assay for the kinase activity in the tumor. This increase in kinase activity was observed to clearly parallel the degree of deregulation of the *mdl* locus, whether that was the result of progressive inactivation or removal of regulating gene(s), or the presence of variable numbers of deregulated *mdl* loci. The specificity of this correlation with the tumor-inducing effect(s) of the *mdl* locus (and illustrating the fortuitous elevation in the brain of the fish bearing crossing-conditioned melanomas) is shown by assaying fish which carry somatically induced melanomas (eg. by carcinogen treatment). In this case the elevated kinase activity is observed only in the tumor tissue, and a parallel increase is not observed in brain tissue from these fish [264a].

These observations raised a number of questions regarding the functional and/or structural relationship between the *mdl* locus of the fish and the *c-src* gene. Anders and coworkers have presented the following possible interpretations [14]: 1) *mdl* and *c-src* expression is coincidental due to a linkage relationship between the two genes, whereby the *c-src* expression is inconsequential for the tumor formation; 2) the *c-src* gene may represent a regulatory gene for the *mdl* locus or vice versa; 3) the *mdl* and *c-src* genes are identical, and 4) the *mdl* locus is composed of a number of tightly linked proto-oncogenes, of which the *c-src* gene is but one. Anders and co-workers have not been able to present definitive data that decides between these possibilities, although Anders currently favors a model based upon possibility 4 [212,14]. In the initial report of Schartl *et al*, a more reasonable suggestion was included, in part based upon the assumption (supported by the data presented here) that the *c-src* gene is a unique gene in the *Xiphophorus* genome, that the elevated kinase activity could be the effect of an

interaction of the *mdl* gene product with pp60^{C-src} or the c-*src* gene itself [263]. This would imply that the *mdl* and c-*src* genes could be on different chromosomes, and that the elevated c-*src* kinase activity may well be playing a role in the tumorigenesis process. However, no evidence could be presented by these authors either supporting or negating this proposal.

The cloning of the c-*src* gene from the *Xiphophorus* genome made it possible to test certain aspects of these models directly. Specifically, the continued backcrossing of the hybrid fish with benign melanoma to the *X. helleri* parent results in effect in the isolation of the *X. maculatus* X-chromosome carrying the *mdl* locus within the *X. helleri* genetic background. If one could distinguish between the *X. maculatus* and *X. helleri* c-*src* genes, then one could directly assay DNA from these highly back-crossed tumor-bearing fish for the *X. maculatus* and/or *X. helleri* c-*src* genes. The presence of the *X. maculatus* gene in these backcross fish would prove linkage between the *mdl* locus and c-*src* (i.e. that the two genes are found on the same chromosome - the X-chromosome), whereas absence of the *X. maculatus* c-*src* gene in these fish would prove that the c-*src* gene is not found on the same chromosome as the *mdl* locus. In addition, in the event that the *X. maculatus* c-*src* gene should indeed map to the same chromosome as *mdl*, the X-chromosome translocation and deletion mutants involving the *mdl* locus could be used in an attempt to map the c-*src* gene relative to the *mdl* locus - i.e. to answer the question whether *mdl* and c-*src* are the same gene or not. As is shown very clearly in Figure 15, the *X. maculatus* c-*src* gene is not found in the backcross fish, proving that this gene is not found on the X-chromosome. In addition, hybridization of the same DNA samples with the v-*src* kinase domain probe under conditions that detect a large number of *src*-related sequences within the genome shows no differences in the hybridizing bands between these DNAs (data not shown), suggesting that no members of the *src* gene family detected by this probe (i.e. specifically the kinase domains of these genes) are found in the vicinity of the *mdl* locus. Therefore all models considered above based upon linkage or identity of the *mdl* locus and c-*src* can be discounted, leaving the original proposal of some form of co-regulation or interaction between the two genes as the most viable option. This result also means that the elevated tyrosine kinase activity detected in the tumor-bearing hybrid fish and which has been shown to be paralleled by a high level of c-*src* mRNA [197] involves the *X. helleri* c-*src* gene. Therefore, any models implicating a role for c-*src* in the melanoma formation or progression must involve an interaction of some sort, direct or indirect, between the introduced *X. maculatus* *mdl* locus and the *X. helleri* c-*src* gene, and this interaction occurs in both the melanoma and in the brain of these fish.

It is conceivable that the *mdl* locus encodes a particular protein product (or products) which normally function in certain derivatives of the neural plate (which gives rise to both neural tube and neural crest) - eg. certain neuron types localized to particular sub-regions of the brain, and certain pigment cell types. For example, basic fibroblast growth factor has been suggested as the natural growth factor for human melanocytes, and it is possible that the *mdl* locus encodes a growth factor receptor (see below). The normal function of the *mdl* gene product may involve an interaction of some sort with the c-*src* gene or its protein product, regulating the amount of pp60^{C-src} and/or its enzymatic activity. That just such an interaction can occur between a growth factor receptor and pp60^{C-src} is shown by the elevated pp60^{C-src} tyrosine

kinase activity that results following the binding of platelet-derived growth factor to its cell surface receptor [113a,244a]. When the *mdl* locus is progressively deregulated, as results from the crossings carried out to generate the melanoma-bearing fish, this interaction with *c-src* (or $\text{pp60}^{\text{c-src}}$) may be intensified. This still leaves open, however, the question whether the *mdl* deregulation alone is responsible for the tumor induction (i.e. in some pathway in addition to and distinct from that involving *c-src*) or whether the elevated (and altered?) *src* tyrosine kinase activity is necessary for the tumor formation.

There has been a report of linkage of a *src*-related sequence to a *mdl*-carrying X-chromosome of *X. maculatus* [317]. In that case the observed band (an approximately 5 kb Eco RI fragment that hybridized with the *v-src* kinase domain probe) that was missing in the DNA sample thought to be a deletion mutant for *mdl* is, based upon the data presented here for the *X. maculatus c-src* gene, not derived from the fish *c-src* gene. The size of the restriction fragment would suggest that this sequence could come from the corresponding genomic locus of clone 20-2 (see Table II). As the *v-src* hybridization probe under the standard stringency conditions employed in this dissertation detects, among a number of other *src*-related sequences, also the 5.0 kb Eco RI band from clone 22-1 (see Fig. 1 and Table II), it must be concluded that the particular deletion mutant investigated here and reported in this thesis did not entail a deletion of the genomic 20-2 locus. The possibility exists that the deletion involved in the mutant used by Vielkind and Dippel [317] is more extensive and includes all or part of the cellular 20-2 locus. Vielkind has recently provided DNA from the same deletion fish as described in their manuscript, and it will be possible to more directly test this possibility using a hybridization probe derived from the 20-2 genomic clone. Nevertheless, that a deletion can occur that inactivates the *mdl* locus but that does not involve the 20-2 locus (strictly speaking, the kinase domain of the 20-2 locus) suggests that even in the event that the 20-2 genomic locus should map to the *X. maculatus* X-chromosome it is not functionally implicated in the tumor induction.

The availability of the cloned *c-src* gene from the fish makes it possible to examine several of the possible relationships between *mdl* and *c-src* regarding co-regulation of the two genes. These would include regulation of expression (potentially of both genes) at the level of DNA methylation, and/or the role of gene amplification/rearrangement.

The analysis of the methylation status of the fish *c-src* gene was performed by way of restriction enzyme digestion of genomic DNA with pairs of enzymes which recognize the same sequence but differ in their sensitivities regarding the methylation of certain bases within that recognition site. The first experiments, utilizing testes DNA, confirmed results reported in a number of other systems that sperm DNA is very highly methylated [320,335]. The very high molecular weight band detected with all probes in the Hpa II digests in Figure 16 are consistent with the majority of Hpa II/Msp I sites within the sperm DNA being methylated at the internal C residue (i.e. $\text{C}^{\text{m}}\text{CGG}$). The ethidium-bromide stained gel (Fig. 16d) confirms the fact that the majority of the CCGG sites in the sperm DNA are methylated at the internal C residue. This result is also consistent with Bird and Taggart's observations concerning whole genome methylation patterns in various groups of animals [29].

The hybridizations using fish melanoma DNA and normal tissues for comparison point towards a potential role for specific demethylation of the *c-src* gene in progression of the tumor. That is, a *c-src* demethylation event has been observed only in a tumor DNA sample derived from highly malignant tumors, whereas the benign tumor samples give no indication of such an event. Although one can consider only the Hpa II digestions, it is apparent from Figure 17c that a specific demethylation event has occurred within the kinase domain of the *c-src* gene that is detectable only in the malignant tumor DNA sample and not in the DNA derived from the benign melanomas. The *Xhomeobox* hybridizations of the Hpa II digests (Figure 17d) show that this demethylation event is specific for the *c-src* gene. Although the elevated *c-src* tyrosine-specific protein kinase activity is detected in both the brain and the melanoma of the tumor-bearing hybrid fish, the *c-src* demethylation event is observed only in the melanoma DNA sample. This suggests that an elevation of *c-src* gene activity can occur in the absence of demethylation of the gene (at least in the domain of the gene detected with the particular probe employed here). This is also supported by the finding of elevated tyrosine kinase activity in the benign melanomas in the apparent absence of a demethylation event within this domain of the *c-src* gene. These results all support a role for the observed demethylation in a later stage of the tumorigenesis - i.e. in the progression of the tumor to true malignancy.

DNA methylation is thought to play a role in the regulation of gene activity by stably altering the local structure of the gene and/or affecting the interaction of trans-acting positive or negative transcription factors usually, but not always, upstream of the structural gene [49,50,72,335]. A number of studies point to a general (but not 100%) correlation between the degree of methylation of a gene and its level of expression - that is, those genes being expressed in a certain cell are hypomethylated in comparison to the same gene in a cell where this gene is not being expressed [49,50,72,335]. This is particularly evident for tissue-specific genes, but not so for house-keeping genes which are generally nonmethylated in all tissues [30]. The best evidence in support of methylation playing an important role in the regulation of eukaryotic gene expression comes from experiments whereby specific genes can be inactivated by experimentally induced methylation [45], and conversely, experiments where genes have been shown to be transcriptionally activated by specific de-methylation [336].

The initial observations supporting a general role for DNA methylation/demethylation in the regulation of gene expression quickly led to models of carcinogenesis based either solely upon, or at least incorporating, DNA methylation [134,135,146,251]. That DNA methylation is an epigenetic event not involving a change in the sequence of the DNA and therefore conceivably reversible was in line with a large body of observations (see Introduction) consistent with a reversible event underlying at least some neoplastic events. The crossing-conditioned melanoma in *Xiphophorus* is also an example of a reversible neoplastic event. First, if the tumor-bearing fish are backcrossed to the *X. maculatus* parent (instead of the *X. helleri* parent), the *mdl* locus becomes progressively regulated (as β *mdl*-containing *X. maculatus* chromosomes are reintroduced), with a return eventually to the normal spotted phenotype [11]. This means that no irreversible germline mutation has occurred in any of the tumor-bearing fish, although this experiment does not discount an irreversible somatic event in the tumor. This is addressed in a second experiment, whereby treatment of the tumor-bearing fish with agents thought to result in

induction of differentiation of the transformed cells results in a degree of reversion to benignity of the tumor [261].

The presentation of these models of carcinogenesis based upon DNA methylation effects led to the search for methylation changes in particular tumors, first at the overall genome level, and then later at the level of particular genes. Although there is some variability in the results [51,59], the general conclusion from these sorts of studies is that hypomethylation of particular genes is often observed in tumor DNA samples [68,84,93,173]. This has also been shown to be the case for several proto-oncogenes, including *c-myc* in hepatoma, leukemia and plasmacytoma [77,236] and *c-ras* [83]. In addition, Feinberg and colleagues report that hypomethylation to the same degree of several genes could be observed in both benign colon polyps and malignant colon carcinomas, suggesting that it is an alteration in the DNA that precedes outright malignancy [99]. However, cause and effect relationships regarding gene demethylation and carcinogenesis are still very unclear. Because the platyfish/swordtail melanoma system presents a tumor that results due to a precisely defined genetic basis, it should be a useful system in which to investigate further the possible role of demethylation of particular genes in this tumor formation process.

The methylation studies presented here suffer, however, from several drawbacks. First, the number of tumors that have been investigated by way of Hpa II/Msp I digestion and hybridization is relatively small, so these results must be regarded as preliminary. Second, the *c-src* hybridizations have only been performed to date with probes from the coding region of the gene, specifically the kinase domain. However, it is to be expected, by analogy to other systems, that some of the methylation changes thought to be playing a role in the regulation of the *c-src* gene may well map to the 5' non-coding sequences of the gene [30,31]. In addition, it has not been possible to date to precisely locate these CCGG sequences implicated in *c-src* demethylation within the available genomic sequence (i.e. it is not known whether they are found in exon or intron sequences). The availability of the cDNA clone will lead to the isolation of the *c-src* 5' regulatory sequences, which will be assayed in a similar manner for specific demethylation events associated with the tumor induction. In addition, the potential role of *c-src* demethylation can be followed during embryogenesis, whereby developmental northern blots have shown a rise in *c-src* mRNA levels during organogenesis [246]. It has been shown in other systems that methylation changes can accompany the developmental activation of a gene [119]. With the sequence becoming available for the *c-src* upstream regulatory sequences it will eventually be possible to precisely locate the Hpa II/Msp I site(s) implicated in demethylation, and possibly to correlate those sites implicated in developmental activation with sites implicated in tumor induction and/or progression.

The third drawback of these methylation studies is that the tumor DNA samples that have been assayed are all derived from tissue pooled from a number of fish. This has the disadvantage that it is not always possible to differentiate between a change that may have occurred in only one or several samples versus a change in all the samples. If indeed demethylation is playing a central role in this genetically precisely-defined tumor system, then one would expect to observe uniform changes in all tumor samples of a particular genotype. If, however, demethylation events are random, then tumors derived from different fish of the same genotype will contain non-uniform methylation changes. Therefore, these experiments need to be repeated in the future with DNA

samples prepared from tumor tissue (as well as brain and liver) from individual carefully-staged tumor-bearing fish.

In addition to the role of regulatory alterations of proto-oncogenes in the causation of various tumors, structural alterations have also been implicated [reviewed in 34]. This could involve: 1) point mutations, whereby a protein product with an enhanced oncogenic potential results [216]; 2) chromosomal translocations that result in the deregulation of the proto-oncogene [47], and 3) oncogene amplification, leading to elevated levels of expression of an often in addition altered gene product [7].

The *Xiphophorus* c-src probes have been used to examine whether amplification and/or rearrangement of the c-src gene plays any role in the tumor formation resulting from hybridization. Analogous to the situation above concerning c-src methylation, there is some indication that a low level amplification (i.e. 3-5x) specifically of the c-src gene is found only in the malignant melanoma DNA sample (see Fig. 18a - compare lanes 1 and 2). Therefore it would appear that the elevated pp60^{c-src} activity detected in the benign melanomas can occur in the absence of any structural alteration of the c-src gene, but that such changes may be playing a role in the progression of the tumor. Figure 18 also presents data consistent with an alteration of the c-src gene and immediate surrounding sequences specifically within the malignant melanoma DNA sample. The faster migration of the hybridizing band in lane 2 of Figure 18a could, however, be due to a gel-running artefact specifically associated with the tumor DNA sample, since it is difficult to purify melanoma DNA totally free of melanin which could conceivably affect the migration of the DNA in the gel. However, the fact that the *Xhomeobox* -hybridizing band detected in lane 2 of Figure 18b shows the exact same molecular weight as the band detected in the normal tissues argues against such a gel-running artefact. Alternatively, the smaller Bam HI band could be the result of a demethylation event in a cryptic recognition site that is methylated in the other tissues in such a way that Bam HI cannot cut. For example, it is known that Bam HI is sensitive to certain methylation events within its recognition sequence - the enzyme will cut GGAT^mC and GG^mATCC but will not cut GGAT^mCC [213].

As pointed out above for the methylation studies, these c-src amplification data suffer from the same drawbacks, and in particular the planned experiments utilizing DNA extracted from individual tumor-bearing fish should make clear whether c-src gene amplification and/or rearrangement is a random process associated with tumor progression or alternatively a component of the genetically defined tumor-formation process. There are a number of reports arguing that the proto-oncogene amplification observed in some systems has a very strong correlation with the progression of that particular tumor [196], and the situation may well be analogous in the *Xiphophorus* melanoma system.

There have been two preliminary reports of oncogene amplification in *Xiphophorus* melanomas from the Anders group [12,15], but for the following reasons these reports are somewhat puzzling. First, the data as presented are very unclear (i.e. the one hybridization that has been presented, performed with the viral src kinase domain probe, is very poor, with a smear of radioactivity that runs throughout the lane that supposedly contains the amplified DNA).

Second, the two reports state that different *src* -related (but interpreted by these workers as deriving from the *c-src* gene) restriction fragments have been highly amplified in each case. In the first report two Eco RI fragments (25 and 0.5 kb) are supposedly amplified greater than 50-fold [12](but there is in fact a signal smear over the entire lane), whereas in the second report it is a 4.9 kb Eco RI band that is supposedly highly amplified [15]. Regarding this last report, it has been shown that another apparent case of 'c-src amplification' in a *Xiphophorus* DNA sample (non-tumorous; prepared in the Genetics Institute, Giessen) involving this 4.9 kb fragment was due to contamination of the genomic DNA with a recombinant plasmid containing a v-*src* insert (precisely, pBR 322 [4.3 kb] plus the Pst I 600 bp kinase domain fragment of v-*src* - data not shown). This result can occur when genomic DNA and plasmid DNA are being prepared without using single-use centrifugation tubes etc, and extreme caution is not exercised. It is highly likely that the case of 'c-src amplification' reported by Anders and co-workers also owes its origin to such a laboratory artefact. The third puzzling aspect of these reports is that Anders and coworkers report (but do not present any of the supporting data) in addition to the 'c-src' amplification also a 'c-yes' amplification (using the viral *yes* probe), and, rather surprisingly, predict that a *c-ras* amplification would also be expected. These proposals are difficult to reconcile with a reasonable molecular model for the tumor induction that occurs in the hybrid fish.

The vast majority of reports in the literature concerning amplified (proto)oncogenes concern genes of the *myc* and *ras* families [see 216 for review]. It is indeed relatively rare that a gene encoding a tyrosine kinase is found amplified in a tumor cell line or in a primary tumor, and in almost all of these cases the genes are of the epidermal growth factor receptor family [216]. Although there is no other substantiated report of a *c-src* amplification in a tumor or tumor cell line, there is one report of a 4.5 x amplification of the human *c-yes* gene in a primary gastric tumor, this being the only reported case of gene amplification of which I am aware involving a member of the *src* sub-family of tyrosine kinases [274]. There may also be some significance to the fact that whereas the *myc* and *ras* genes can be found amplified to very high degrees (eg. over 500-fold for N-*myc* in certain neuroblastomas and approximately 60-fold for Ki-*ras* in the Y1 mouse adrenocortical tumor), the cytoplasmic tyrosine kinases (i.e. the one report of *c-yes* amplification in a primary tumor and one report of *c-abl* amplification in a chronic myelocytic leukemia cell line) have only been found to date amplified to only minor degrees (eg. 4-8 fold). The several reports of amplification of the epidermal growth factor receptor gene include instances of both low level amplification (eg. 2-4 fold) as well as cases of greater amplification (eg. 30-100 fold) [216]. This may relate to the fact that one method (of several) of modifying the response of a particular cell to a certain growth stimulus is to simply increase the number of receptors at the cell surface for that particular growth factor, with, if the growth factor is not in limiting supply, a growth response of the cell that is a function of the number of receptor molecules. The cytoplasmic tyrosine kinases however are enzymes that can potentially operate independently of extracellular growth factors and, as with most intracellular enzymes, only a small number of protein molecules is sufficient to bring about the physiological response within the cell. In this case a slight change in substrate specificity could have a much more drastic effect within the cell than simple increase in number of enzyme molecules.

The original question behind the experiments discussed in this section of the dissertation was the basis of the co-regulation between the *mdl* locus and the *c-src* gene in the *Xiphophorus* melanoma system. Although some insights have been obtained, a more direct approach would involve the analysis of the *c-src* gene from the tumor DNA, as well as the cloning of the *mdl* locus. The cloning of the *c-src* gene from the tumor would be relatively straight-forward, and would enable a direct determination of whether any structural changes of the gene accompany tumor formation and progression. The precise characterization of the methylation status of the *c-src* gene in the tumor could be achieved through use of genomic sequencing strategies. The cloning of the *mdl* locus is presently underway in this laboratory, and is being approached in two ways, both of which are complementary to one another. First, a bioassay exists for the deregulated *mdl* locus, whereby DNA from a *X. maculatus* strain containing one or several deregulated *mdl* loci is injected into the neural crest region of a *X. helleri* embryo [316]. The precursor cells of the melanocytes and melanophores derive from the neural crest [141], and if one or more of these precursor cells takes up the deregulated *mdl* locus this can lead to the appearance of a clone of macromelanophores in the skin of the *X. helleri* embryo (or young fish). Calculations of the size of the DNA reaching the neural crest region suggested that the functional *mdl* locus could conceivably be contained within a fragment of DNA of 40-60 kb, and it was therefore considered feasible that a complete cosmid library of the *X. maculatus* genome would contain a number of functional *mdl* clones. By applying the process of sib-selection [see 189 for an example of sib-selection applied to cloning from gene libraries] using the embryo injection bioassay it would be theoretically possible to clone out the functional *mdl* locus. A large cosmid library from a *X. maculatus* strain containing two deregulated *mdl* loci has been prepared, and some initial characterization of the library has been performed (i.e. determination of total library size; calculation of average insert size etc. - data not shown) in preparation for embryo injection. In addition, a large number of embryo injections with control plasmid DNA have been conducted in order to perfect the technique as well as to monitor the uptake and maintenance of the exogenous DNA (data not shown). However, before undertaking injection of the embryos with the library DNA, it was deemed best to coordinate this approach with the approaches outlined briefly below.

Modern reverse genetics, as applied to the cloning of a number of human genetic disease loci [225], can be applied to the problem of isolating the *mdl* locus in three general ways: 1) the isolation of RFLP's which map to the vicinity of the *mdl* locus and which can be physically mapped relative to one another by pulsed-field gel electrophoresis; 2) the generation of subtraction libraries [171] which make use of chromosomal deletions involving the *mdl* locus to clone specifically those sequences involved in the deletion; and 3) the generation of *X. maculatus* X-chromosome specific genomic libraries (in the *X. helleri* genetic background *X. maculatus*-specific clones would be isolated with the aid of a dispersed *X. maculatus* - specific repetitive DNA sequence). All of these approaches are presently underway in the laboratory, and all have given positive results to date. Specifically, a *v-erb* B-related RFLP has been identified, cloned and partially sequenced that maps very near to the *mdl* locus (indeed, could derive from the *mdl* locus itself) and which has certain attributes of a gene of the receptor/tyrosine kinase family [2,265; J. Wittbrodt, unpublished; D. Adam, unpublished]; a library has been prepared covering a deletion

thought to include the *mdl* locus and individual clones from this library are being screened in the hope of identifying clones mapping to the X-chromosome and more specifically to the region of the *mdl* locus [330]; and *X. maculatus* repetitive DNA clones have been isolated that enable the distinction between *X. maculatus* and *X. helleri* sequences [324]. Therefore the stage is set whereby one will very soon be able to prescreen particular libraries for those clones that map to the region of the *X. maculatus* X-chromosome containing the *mdl* locus, and then inject this pre-selected group of clones into the *X. helleri* embryos. Because the embryo injection is extremely work intensive, this pre-screening of clones prior to injection will lead to a major reduction in the man-hours involved in the sib-selection of the clones and the isolation of the functional *mdl* locus. The original embryo injection experiments, performed with native total DNA indicated, due to their strict co-transfer [316], a very tight linkage between *mdl* and *Rmdl* (consistent with the proposal that in effect the *Rmdl* locus comprises the non-coding 5' regulatory sequences for *mdl*). Therefore it is to be expected that the *mdl* cloning will also lead to the characterization of the various linked *Rmdl* alleles (i.e. R^{Sd} , R^{Sp} , R^{Sr} , etc.), with an expected insight into the molecular basis for this particular allelic series.

IV. DISCUSSION

3. EVOLUTIONARY ASPECTS OF THE SRC/TK GENE FAMILY

If genes have evolved predominantly from exon-shuffling and intron deletion, then the positions of introns should portray the evolutionary history of a gene, as has indeed been demonstrated for a number of genes [35,52,96,97,190,198,237,304]. As is shown in Figure 13, the *Xiphophorus* *c-src* gene shares the same exon-intron configuration as found in the *src* gene-family members of higher vertebrates. Figure 28 illustrates that this exon/intron configuration is specific for the vertebrate *src* gene-family, and is not shared by other component gene-families of the tyrosine kinase super gene-family. Specifically, the human *c-fes/lps* and the mouse *c-abl* genes have exon/intron configurations that are completely different to that found in the *src* gene family (Figure 28a). One exon/intron junction in the *c-fes/lps* gene is moved by only one codon relative to a junction in the human *c-src* gene, which could be interpreted as resulting from a process referred to as 'intron-sliding' [190]. However, the fact that both genes are made up of a relatively large number of small exons raises significantly the chance that such an observation would result purely by coincidence. In addition, the fact that a close correlation between exon/intron junctions between these two genes is only seen in one case also argues against any obvious evolutionary relationship by descent between these two gene families.

Figure 28b also serves to illustrate the disparity between the exon/intron arrangement of the vertebrate *src* gene family members and the *src*-related genes isolated from the arthropod *Drosophila melanogaster*. The *src*-related gene *Dsrc* 64B (located at position 64B on chromosome III) is commonly referred to as the *Drosophila* *c-src* gene, since it displays the highest sequence similarity to the vertebrate *c-src* genes. However, this is actually a misnomer, in that this gene displays as much sequence similarity to the human *c-yes* gene as it does to human *c-src* [283]. The second *Drosophila* '*c-src*' gene, *Dsrc* 28C (mapping to position 28C on chromosome II), is also equally related to *src* and *yes*, although the level of this similarity is reduced relative to *Dsrc* 64B [118]. There are in addition several other *Drosophila* genes that have been isolated and been shown to display some sequence relationship to *c-src*, in these cases due to the fact that these genes also encode tyrosine kinases [120,121,180; reviewed in 279]. Figure 28b demonstrates that *Dsrc* 64B and *Dsrc* 28C share no obvious relationship to one another at the level of exon/intron arrangement, and in addition both genes share no apparent relationship to the vertebrate exon/intron conformation. The only possible exception is the intron separating exons 4 and 5 in the vertebrate *src* gene family, which coincides exactly with one of the introns of *Dsrc* 28C. The significance of this observation is difficult to measure, in large part again due to the absolute noncoincidence of exon/intron junctions throughout the rest of the three genes. It could, however, indicate a relationship by descent between *Dsrc* 28C at least and the vertebrate *src* gene family.

If this were the case, then, by analogy to the situation for the triosephosphate isomerase gene and a number of other genes (illustrating gene evolution through intron loss) [35,52,96,97,190,198,237,304], the ancestral gene for the *src* gene family must have had an extremely complex exon/intron conformation. According to this model, the ancestral gene

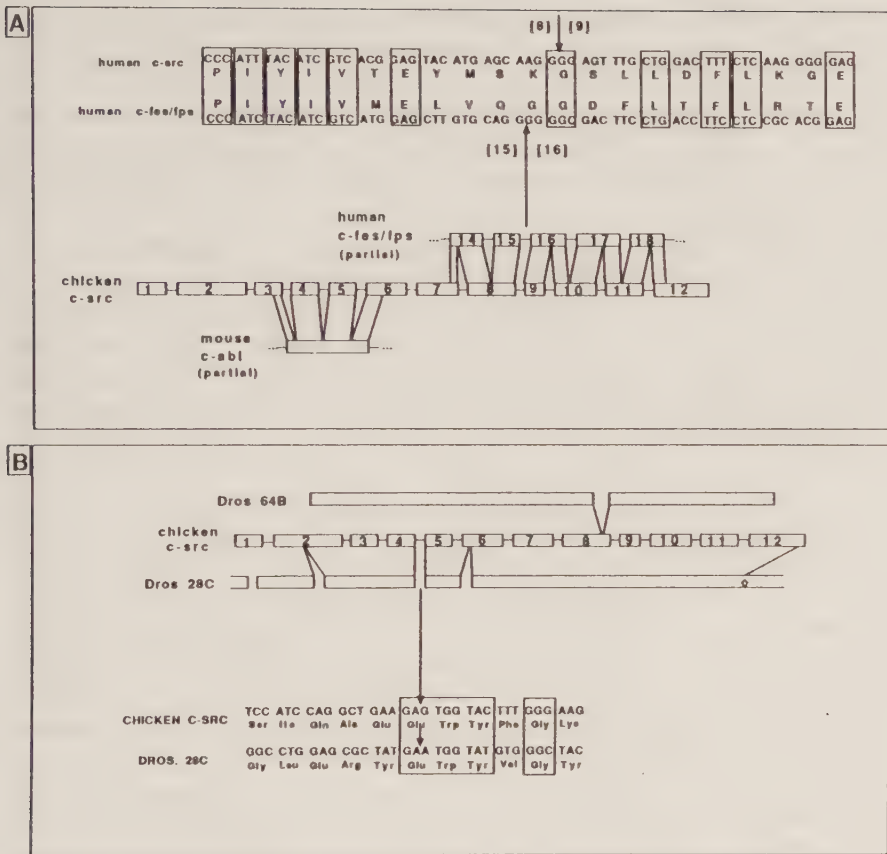


FIGURE 28

src-family exon/intron configuration

A) Chicken c-src exon/intron configuration compared to partial genomic sequences from the mouse c-abl and the human c-fes/lps genes.

B) Chicken c-src exon/intron configuration compared to the Drosophila Dsrc64B and Dsrc28C genes.

duplicated within the lineage that would give rise to the modern-day arthropods following the vertebrate/arthropod divergence, giving rise to *Dsrc* 64B and *Dsrc* 28C. In addition, a large degree of intron loss, particularly in the *Dsrc* 64B gene, gave rise to the present-day *Drosophila* genes. In the other lineage (i.e. leading to the vertebrates), intron loss must also have occurred, giving rise to the ancestral gene for the vertebrate *src* gene-family. This loss must have occurred prior to any of the gene duplication events that would give rise to the members of the present-day gene family, and this exon/intron conformation has been maintained stably ever since, as all members of the vertebrate gene family possess it. This would mean then that all these gene duplication events must have occurred after the vertebrate/invertebrate divergence (normally placed at around 700 Myr ago) and that a single progenitor sequence gave rise to the vertebrate *src* gene family.

Alternatively, the invertebrate and vertebrate *src* genes may not be directly related by descent, and instead illustrate the phenomenon of convergent evolution. This problem may only be approachable through the examination of the genomic arrangement of *src* -related genes in modern-day organisms thought to approximate common ancestors for both the vertebrate and invertebrate lineages, particularly of those organisms thought to approximate the last common ancestor for the lineages. This will require an extension of our current knowledge concerning this divergence event.

When a phylogenetic (i.e. least parsimonious) analysis is performed upon the amino acid sequences of the to-date characterized members of the *src* gene-family (human *c-src*, *c-yes*, *c-lyn*, *c-fgr*, *c-hck*, *c-lyn*, mouse *c-lck*, and chicken *c-tkl*), it is found that the gene family is composed of two major branches (*src/yes/lyn/fgr* and *lck/tkl/hck/lyn* - see Figure 29). Because *c-src* and *c-yes*, and *c-lyn* and *c-tkl*, are the two most closely related pairs of this gene family, this would suggest (but does not prove) that the *src/yes* and the *lyn/tkl* duplications were the most recent, and therefore that all other duplications within the gene family probably occurred prior to this time. The data presented here show that the *c-src* and *c-yes* genes are evolving at approximately the same rate. If that rate of gene divergence is assumed to be valid for all gene family members (which may not be unreasonable as all family members encode for proteins with the same enzyme activity, although they may well differ in their substrate specificity), then it is possible to draw an evolutionary tree for the *src* gene family in which all gene duplication events are placed relative to an absolute time scale. Such a tree is presented in Figure 29. These calculations suggest that the original split of the gene family into two major branches was an extremely ancient event. In addition, as is seen from Figure 29 (see also Appendix, Fig. 4), *c-lck* and *c-tkl* are approximately as related to one another as *c-src* and *c-yes* are to one another. Therefore the *tkl/lck* duplication may have occurred at around the same time as the *src/yes* duplication, and could well derive from the same theoretical genome duplication event.

Figure 29 is very tentative due to the assumption of a constant divergence rate for all gene family members equal to that determined for *c-src* and *c-yes*. Until recently there was no sequence available from different organisms for any other members of the gene family, so it was not possible to attempt to confirm this assumption. However, the recent publication of the

sequence for the mouse *c-hck* gene [159] allows a comparison to the published human sequence and the calculation of a divergence rate for this gene. Surprisingly, the derived divergence rate is significantly higher than that determined for either the *c-src* or the *c-yes* genes. For the *c-hck* gene the %PAM/100 Myr and UEP values (at the level of amino acids) are 8.86 and 11.3 respectively, whereas for the *c-src* gene (based, as for *c-hck*, on mouse/human comparisons) the values are 1.23 and 81.4 respectively. That is, the human and mouse *c-hck* genes encode proteins that are ~5.8X more diverged than the respective *c-src* proteins. Therefore the gene duplication giving rise to the *c-hck* gene may well be placed too early in Figure 29. It is conceivable, furthermore, that the gene duplication event that gave rise to *c-hck* may also be the result of one of the hypothetical genome duplication involving *c-src*, *c-yes*, *c-lck* and *c-trl*. A more accurate (i.e. robust) tree must await additional sequence data for gene family members from more organisms. The indications are that the vertebrate *src* gene family arose from one single progenitor sequence following the divergence of the arthropods - i.e. that all *src* family gene duplication events have occurred within the last 600 Myr or so. Therefore the assumption of constant evolutionary rate for all family members is apparently not valid (as has been shown for *c-hck*), and it is furthermore conceivable that much of the gene family diversity is indeed the result of several whole genome duplications. Figure 29 presents the correct topology for the *src* gene family evolutionary tree, but the only duplication event placed with any certainty against the absolute time scale is that for *src* and *yes*.

In addition to tracing changes in the exon/intron arrangement of members of a particular gene family, it is also possible to correlate more minor changes within one or more particular exons with evolutionary events within the gene family. Within the *src* gene family there is a small degree of variability within the exon/intron framework shared by the different members. This is most pronounced in exon 2, encoding for the extreme NH₂-terminus of the respective protein products, but is not limited to this exon. Variations that are found conserved in several members of the gene family are most readily accounted for by assuming that that particular sequence motif was present in the progenitor sequence that gave rise to these family members. An example of such a motif is found in *lyn*, *hck*, *trl* and *lck*, where in all four genes the same amino acid residues are missing from exon 3 (1 residue), exon 7 (1 residue) and exon 12 (2 adjacent residues) relative to the human *c-src* gene. This is strong circumstantial evidence for a single common progenitor for these four genes, supporting the proposal that the *src* gene family was split early into two branches (see Fig. 29).

There are also changes present in only one member of the gene family, which are most readily explained as the result of changes specifically within that gene sometime following its emergence. An example of such a sequence variation is a block of seven consecutive extra amino acids in exon 2 of the human *c-yes* gene. The complete characterization of the members of the fish *src* gene family, as well as the corresponding genes from chicken, mouse, man and other relevant vertebrates will allow a more precise definition and timing of these sub-family specific changes. In particular, the sequence of the mouse, chicken and fish exon 2 from *c-yes* will allow a crude timing of the appearance of these seven amino acids. As a second example, three identical extra amino acids are found in exon 2 of the human *c-src* gene, as well as in the mouse *c-src* gene, but are not present in the chicken *c-src* gene. The most logical explanation for this observation is

that sometime following the divergence of the birds and prior to the mouse-human divergence these three extra amino acids arose in the mouse-human progenitor *c-src* sequence. Alternatively, these amino acids may have been present in the *c-src* sequence prior to the avian divergence, but have been specifically lost from the *c-src* gene within the avian lineage. If also present prior to the fish divergence, then the fish *c-src* gene may encode a protein which contains these residues. Again, complete exon 2 sequence from the fish *c-src* gene could help in distinguishing between these two possibilities. It is conceivable that these intra-family variations play a role in the particular gene-specific function(s) which are assumed for the various gene family members.

The data presented in this thesis allowed the calculation of the rates of evolutionary divergence for the *c-src* and *c-yes* genes. The calculated rates at both the nucleic acid level (3.0 and 3.6 changes per 100 residues per 100 Myr for *c-src* and *c-yes* respectively) and amino acid level (1.1 and 0.9 changes respectively) were significant for two reasons: 1) the divergence rates are very similar, implying that since the *src/yes* gene duplication both genes have been evolving at the same rate (i.e. that both genes are under very similar selective pressure); and 2) the divergence rates are very low (see Table VIII). That is, both the *c-src* and *c-yes* genes and their respective protein products are very highly conserved over evolutionary time. This is demonstrated by the very high UEP values in Table VIII. In fact, as is shown in Table VIII, the cellular oncogenes as a whole encode for very highly conserved protein products. This is consistent with their proposed central role in the control of cellular proliferation and/or differentiation.

Gojobori has also reported divergence rate values for the *c-src* gene, which are however about two times higher than the values reported here [102,103]. This is probably due at least in part to the fact that Gojobori compared only chicken and *Drosophila c-src* (*Dsrc* 64B) sequences, which entails two potential problems. First, as pointed out above in the discussion concerning *c-src* exon/intron configurations, one can not be sure that the *Drosophila src*-related sequence mapping to 64B on the polytene chromosomes is truly the orthologous (i.e. homologous by descent) gene to the vertebrate/mammalian *c-src* gene. Second, a number of groups, looking at several quite different genes, have shown that the lineage leading to *Drosophila* is one of several 'fast-clock' lineages - i.e. that the rate of gene divergence within this lineage is higher than that found within most other lineages for those particular genes [42,149]. This would then certainly lead to the calculation of a higher gene divergence rate when only chicken and *Drosophila* sequences are compared. In the present work only sequences were used for divergence rate calculations where, based upon gene structure considerations, one could be sure that purely orthologous genes were being compared.

Consistent with the observations from a number of groups investigating other genes [42,332], it is also reported here that for the *c-src* gene the lineage leading to modern-day rodents is apparently a 'fast-clock' lineage. Specifically, assuming a human/mouse divergence time of 70 Myr, the mouse *c-src* gene is approximately three times more diverged at the nucleic acid level than expected, based upon the other comparisons. This is particularly intriguing, given that at the amino acid level the mouse pp60^{*c-src*} protein is extremely highly conserved (see

TABLE VIII

Rates of protein evolution.

PROTEIN	UNIT EVOLUTIONARY PERIOD ^a	PROTEIN	UNIT EVOLUTIONARY PERIOD ^a
Histones		Oxygen-binding proteins	
H4	400	Myoglobin	6
H3	330	Hemoglobin (alpha)	3.7
H2A	60	Hemoglobin (beta)	3.3
H2B	60	Secreted enzymes	
H1	8	Trypsinogen	6
Fibrous proteins		Lysozyme	2.5
Collagen	36	Ribonuclease	2.3
Crystallin	22	Immunoglobulins	
Intracellular enzymes		- chains (C)	1.7
Glutamate dehydrogenase	55	- chains (V)	0.8
Triosephosphate isomerase	19	Other proteins	
Carbonic anhydrase B	4	Albumin	3
Electron carriers		Lactalbumin	2.3
Cytochrome C	15	Fibrinopeptide A	1.7
Ferredoxin	6		
Hormones		Oncogenes ^b	
Glucagon	43	c-ras	100
Insulin	14	c-src	40 ^c
Growth hormone	4 ^d	c-myc	10.5

a) Unit Evolutionary Period is the average time, in millions of years, required for a 1% difference in amino acid sequence to arise between two lineages. [From Wilson, A.C. et al. Ann. Rev. Biochem., 1977]

b) All values based upon fish/human comparison, assuming divergence 400 Myr ago.

c) Based upon other comparisons, an average UEP of approx. 100 Myr is reached.
[mouse/human = 160; mouse/chicken = 100; human/chicken = 120]

d) Fish/human, fish/bovine and fish/rat comparisons give an UEP value of 6.3

Table V; mouse-human comparison). Therefore a number of silent third-base changes have occurred in the mouse *c-src* gene, very few of which have led to an amino acid substitution. It would seem therefore that the faster divergence rate at the nucleic acid level is not due to any reduction of the selective pressure upon the *c-src* protein product. The suggestion has been made that the increased divergence observed in the mouse at the DNA level may be the result of differences between the human and the mouse regarding capabilities of DNA repair [42]. If this were so, this would seem to imply that the chicken and the fish, for example, do not share with the mouse this reduced capacity for DNA repair. Wilson has questioned results suggesting 'fast-clock' lineages, mainly due to the uncertainties present in the assumed divergence times based upon paleontological evidence [220,329]. But the distinction apparent here for the *c-src* gene between divergence at the DNA and amino acid levels within the same lineage would argue for an actual rate change, independent of divergence time uncertainties. Indeed, there are some indications that suggest that the rate of evolution of the pp60^{C-src} protein has been decreasing over evolutionary time - all comparisons involving the fish *src* amino acid sequence lead to the highest divergence rate calculations whereas the most recent comparison (mouse-human) gives rise to the lowest divergence rate calculation. This change of divergence rate at the amino acid level has occurred although the divergence rate at the nucleic acid level has remained somewhat steady (or has in fact risen, as in the case of the mouse). These data support the notion of a constant molecular evolutionary clock at the DNA level for the *c-src* gene (although perhaps not applicable to all lineages), while at the same time suggesting that the *c-src* protein has been evolving under selective pressure towards a relative optimum, whereby very few residue substitutions can now be tolerated following the human-mouse divergence approximately 70 Myr ago.

It is interesting to note that the pp60^{C-src} protein contains a very highly conserved kinase domain, as well as a domain at the NH₂-terminus that is not so highly conserved, either between gene family members within one organism, or even between *c-src* genes in different organisms. Therefore, this protein (and the encoding gene) can give rise to sequence data useful for comparisons over large evolutionary distances (using the kinase domain data as presented here) as well as comparisons over shorter evolutionary distances (i.e. times) using the faster diverging NH₂-terminus sequence data.

However, one must keep in mind the uncertainties and assumptions associated with these sorts of analyses. Obviously the assumed divergence times are very approximate, and when based upon very incomplete fossil records could be severely in error [328,329]. In addition, it has been shown for other multi-gene families that when the sequence comparisons become sufficiently detailed it is possible to detect gene conversion events [187]. These gene conversion events in effect extend the sequence similarities between certain members within the family, with the result that both lower divergence rates and closer relationships for certain family members would be calculated than is actually the case due simply to point mutations. Because the level of sequence comparisons within the *src* gene family have not yet reached this level of sophistication (gene conversion events are difficult to detect over greater evolutionary distances), these effects have not been taken into account here.

Lastly, the assumption had been made (with the exception of the mouse lineage; see above)

that both *c-src* and *c-yes* have been evolving at the same constant rate over evolutionary time. The comparisons from the human, chicken and fish sequences supported this assumption. Because sequence was only available for the *c-yes* gene from fish and human, no independent check of this assumption could be performed for the *c-yes* gene. However, when one gene duplicates within a genome, it would be predicted that one copy of the gene could continue to produce the original gene product (presumably under precisely the same selective pressure as before) whereas the new gene copy would at least initially be free to diverge at an increased rate. This new gene may eventually become inactivated or alternatively may acquire a new function - and therefore come again under selective pressure. How quickly the new gene copy acquires a new function would determine how long the gene had been basically 'drifting' - i.e. mutating/diverging at an increased rate. This situation is highlighted schematically in Figure 30 for two genes that following the divergence event apparently are diverging at the same rate (determined from the available sequences) - i.e. the apparent situation for *c-src* and *c-yes*. It can be seen from Figure 30 that this could lead to an overestimation of the time of the gene duplication event. The magnitude of this overestimation would be a function of both the time required for the second gene to come under selective pressure (i.e. the time required to acquire a selectable function) and the 'drift' divergence rate within this time period. One must therefore be aware that the time for the *src/yes* divergence presented here (550 Myr ago) and discussed further below is an approximation based upon available evidence. It is interesting to note in this context, however, that Wilson and coworkers, in analyzing the globin and the lactalbumin/lysozyme gene duplications, found the evidence for an increased evolutionary rate in one of the recently duplicated genes equivocal [328].

Based upon the calculated divergence rates for the *c-src* and *c-yes* genes and the determination of the genetic distances between the *c-src* and *c-yes* genes in both fish and in humans, the *src/yes* duplication was estimated to have occurred approximately 550 Myr ago. This calculation carries with it some potentially interesting evolutionary implications. First, it places the *src/yes* duplication after the divergence of the arthropod lineage and at the time of the emergence of the first proto-chordates. Second, the early Cambrian (the Cambrian is a geologic interval that contains rocks formed between about 570 and 505 Myr ago) was a period of intense evolutionary activity, whereby fossil representatives of practically all major phyla can be recovered from early Cambrian deposits [199]. Third, it is consistent with the suggestion by Doolittle, based predominantly upon analyses of plasma proteins, that as a result of gene duplications a burst of new proteins emerged some 500-600 million years ago [74]. This would place these gene duplication events just before the appearance of vertebrates in the fossil record. Fourth, as Ohno has most strongly argued, major evolutionary steps may be the direct result of whole genome duplications [222], whereby one copy of each gene is freed from selective pressure, increasing the possibility of the emergence of major variations on a developmental and/or structural theme. This evolution by genome duplication model has been presented in conjunction with the early evolutionary history of the phylum Chordata, whereby a genome duplication event was suggested as underlying the evolution from a tunicate-like organism (Urochordata) to an organism resembling the present day *Amphioxus* (Cephalochordata), as well as underlying the

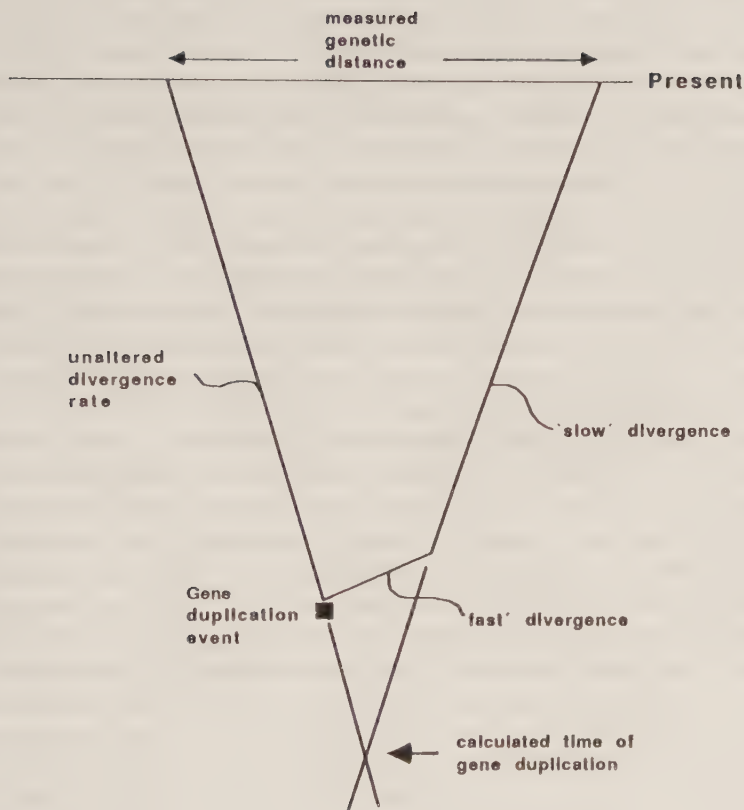


FIGURE 30

Schematic presentation of the possible overestimation of the time of a gene duplication event due to a theoretical increased divergence rate in one of the duplicated sequences immediately following the duplication. The gene to the left continues to make the original gene product and has therefore remained under the selective pressure that had been operating on this protein. The duplicated gene to the right is initially free of selective pressure and can therefore, until acquiring a 'new' function, diverge at an increased rate.

next step from this *Amphioxus* -like organism to an organism thought to resemble the present-day jawless fishes (eg. lampreys)[224]. The proposed *src* /*yes* duplication at 550 Myr ago could well be the result of a genome duplication of this sort, and therefore could be consistent with Ohno's hypothesis. In addition, the gene duplication events proposed by Doolittle may also be the result of whole genome duplication events occurring in the early chordate evolution. The emergence of tunicate- and *Amphioxus* -like organisms is believed to have occurred approximately 500-550 Myr ago. By looking closer at the *src* -related sequences within the modern-day tunicate and *Amphioxus* genomes, it may be possible to more directly test this hypothesis.

As a first approach towards such an analysis genomic Southern blots from tunicate and *Amphioxus* DNA were hybridized with v-*src* and v-*yes* specific probes in an attempt to ascertain whether independent *src* - and *yes* -related sequences were present in these genomes. The results, although relatively crude, are consistent with the presence in the genome of the earliest proto-chordates of separate *src* - and *yes* -related sequences, supporting the proposed *src* /*yes* duplication time. The apparent lack of v-*yes* reactive sequences in the lamprey genome may reflect specific loss of the c-*yes* gene in the lineage leading to modern-day lampreys, or, alternatively, a high sequence divergence rate specifically within the c-*yes* gene such that under the stringency conditions employed no hybridization was detected. The hybridizations with DNA prepared from organisms whose lineages are thought to have diverged before the proposed gene duplication event are also consistent with the proposed duplication time: in the crab and the tapeworm, only a single *src* -related band is detected in the genomic DNA and no *yes* -related band has yet been detected. These results suggest that under the stringency conditions employed only one *src* -related gene is detected in these organisms, but no separate c-*yes* gene - consistent with both the proposed duplication date as well as the ancient evolutionary origin of the c-*src* gene (see below). The results with the *Drosophila* genomic Southern blots are again consistent: the several bands detected with the v-*src* probe probably indicate some of the several *src* -related/tyrosine kinase genes that have been cloned from *Drosophila* [118,120,121,180,279,283], and the single v-*yes* reactive band detected probably represents *Dsrc* 64B, which is as related to the v *yes* probe as the v-*src* probe and is the most *yes* -related sequence in the *Drosophila* genome [283]. *Dsrc* 28C was probably not detected under these stringency conditions of hybridization and washing because, although also related to both probes the level of similarity at the nucleic acid level is reduced relative to *Dsrc* 64B.

The next step would be to prepare genomic libraries from several of these key organisms in vertebrate evolution (eg. tunicate and *Amphioxus*) and isolate and characterize these *src* - and *yes* -related sequences. The same initial questions would arise as with the characterization of the *src*-related clones isolated from the *Xiphophorus* library: what is the genomic arrangement of the genes within the genome (i.e. the exon/intron configuration) and what is the extent of the diversity of the *src* gene-family within these organisms. Would it be possible to trace a gene (or genome) duplication event within this gene family by looking close enough at genomic sequences of extant organisms?

The observation by Scharff and Barnekow of a *src* -related tyrosine kinase activity in the sponge [262], as well as *src*-related sequences in sponge DNA and mRNA [222], suggested that

the *c-src* gene was of an extremely ancient origin. The data presented above concerning the timing of the duplication events within the *src* gene family are also consistent with this. Presently, *src*-related sequences from both the fresh-water sponge *Spongilla lacustris* and the coelenterates *Hydra attenuata* and *Hydra viridissima* are being characterized in our laboratory.

Hydra is the first organism in phylogenesis to contain nerve cells, and the animal also contains a very rudimentary system of interconnecting nerve cells referred to as a 'nerve net' [37,287]. As the *c-src* gene is found to be preferentially expressed in certain neural tissues in higher vertebrates and mammals [56,90,262,285] (as well as in invertebrates - eg. *Drosophila* [283]), it was hoped that *Hydra* might serve as a simpler model system in which to study this nerve cell-restricted expression of the *c-src* gene. Tyrosine kinase assays (using antisera from rabbits bearing Rous sarcoma virus-induced tumors) indicate that the *Hydra c-src* gene is indeed expressed preferentially in the nerve cells of the animal (Schartl, M., Holstein, T., Robertson, S., David, C. N. and Barnekow, A., in preparation). First, the head of the animal (tentacles and hypostome) is enriched in both cells of the nerve lineage (nerve cells and nematocytes) and the *src*-related tyrosine kinase activity. Second, in mutant animals depleted of cells of the nerve cell lineage, the *src* tyrosine kinase activity is reduced in parallel with the degree of nerve cell depletion. Third, purified nematocytes are negative for the kinase activity. These results lead one to conclude that the bulk of the observed tyrosine kinase activity reactive with the rabbit antisera is derived from the *Hydra* nerve cells.

The Southern blots of *Hydra* DNA hybridized with the *v-src* and the fish *c-src* kinase domain probes demonstrate quite clearly the detection of *src*-related sequences within the *Hydra* genome (Figure 27). Experiments aimed at the cloning of the *src*-related sequences from *Hydra viridissima* are currently in progress. The genomic *c-src* clone should allow the isolation of sub-fragments that can be used to probe the expression of the *c-src* gene at the mRNA level in various *Hydra* cell types as well as during developmental processes. For example, 1) *Hydra* cells could be physically sorted and each cellular fraction assayed for expression of the *c-src* gene; 2) the hydra (~ 50,000 cells) can be separated into single cells (the cells maintain their characteristic morphologies), attached to a surface, fixed, and then assayed by *in situ* hybridization to determine which cells or cell type(s) are expressing the *c-src* gene; or 3) the entire animal can be fixed, sectioned and assayed by *in situ* hybridization as above. As the indications are that the *c-src* gene product in *Hydra* is found preferentially in nerve cells [Schartl, M. et al. in preparation], these experiments outlined above could be done in conjunction with immunohistochemistry, utilizing a collection of monoclonal antibodies that react with specific subsets of *Hydra* nerve cells [32]. This would be a direct way to determine whether all nerve cells in *Hydra* express the *c-src* gene at high levels of or only specific subsets. In higher vertebrates it clearly appears that only certain nerve cells are expressing elevated levels of *c-src* mRNA [246]. In addition, these same sorts of studies could also be performed in *Hydra* undergoing specific developmental processes that have long been studied in this organism: eg. regeneration and budding [43,71].

The sponge is the most primitive metazoan, but nevertheless a relatively complex organism. There has been some debate for a number of decades as to whether the sponge also

contains nerve cells [147,174,186,233,239]. Cells resembling nerve cells have on occasion been described in sponges [238,240], and there has been histochemical identification of some vertebrate-type neurotransmitters in sponges [176]. However, a legitimate action potential has not been demonstrated in sponges [174,175,240], and the slower responses observed in sponge have been attributed to a conductive epithelium (as is also found in *Hydra*) and not to a classical nervous response [174,240]. Nevertheless, the sponge is made up of a number of different cell types and a coordinated response is possible - therefore intercellular communication plays a very key role in the coordination of the sponge [181]. In particular, gap junctions are thought to be central to the conduction through the epithelium (again, as in *Hydra*). Some preliminary tyrosine kinase assays have been performed on purified cell fractions of sponge (Barnekow, A and Scharlt, M., unpublished), where the bulk of the tyrosine kinase activity was found associated with the epithelial cell fraction. However, this fraction was contaminated to a certain degree with other cell types and therefore these results are not yet definitive. Pure fractions of the archeocytes and choanocytes were on the other hand tyrosine kinase negative. As outlined above for *Hydra*, the sponge *c-src* gene will serve as a probe which by way of *in situ* hybridization to either cell fractions or sections will determine which particular sponge cells express the *c-src* gene and when. In addition, the sponge is also well suited for the analysis of certain developmental processes, such as the hatching of gemmulae and the early development and growth of the sponge [245], as well as the classic phenomena of cell aggregation which have been studied in sponges for many years [92,205,313]. The three sponge *src*-related clones are presently undergoing further characterization and sequencing by S. Otitile in our laboratory as part of her ongoing Ph.D thesis. Partial sequence has been obtained from sponge clone 4-1 consistent with this sequence deriving from a *src*-related sequence from the sponge genome (S. Otitile, unpublished).

Several members of the *src* gene family in present-day vertebrates or mammals are expressed exclusively in cells of hematopoietic origin (eg. *lck*, *hck*) [191,243,338]. This raises the possibility that particular gene duplication events within this gene family may have been associated with certain steps in the evolution of the immune system. This could be more directly tested by characterizing gene family members and in particular their expression patterns in cells and tissues from organisms selected as representing particular stages in the evolution of the immune system - eg. the amoeboid mesenchymal cell of sponges and coelenterates, protostome coelomocytes and haemocytes, and, more importantly, the first lymphocyte-like cells in some echinoderms (axial organs) and tunicates (gill-associated lymph nodules) [254].

What might the function(s) be of the *c-src* gene? First, it is quite clear that a high level of expression of the *c-src* gene in vertebrates is found in (but is not restricted to) certain nerve cell types [56,90,262,285]. Therefore the gene apparently does not have a function central to the physiology of all nerve cells. Second, analyses conducted on purified synaptic vesicles have demonstrated significant levels of tyrosine kinase activity [231]. Some very preliminary results (M. Scharlt, unpublished) indicate *src*-related tyrosine kinase activity associated with highly purified synaptosomal vesicles of rat brain. Therefore, pp60^{C-SRC} may be playing a role in certain

types of synaptical transmission. This is consistent with the *in situ* localization of c-src mRNA [246] and immunoreactivity [285] in the adult retina, where pp60^{C-SRC} could conceivably be playing a role in the predominantly uni-directional signal transduction from photo-receptor cells to the optic nerve. These adult nerve cells are truly post-mitotic cells, indicating that the elevated c-src expression is not associated with cellular proliferation. The elevated c-src expression observed during embryogenesis also occurs following the bulk of the cellular proliferation, and is correlated more with a number of differentiation events occurring during organogenesis [264]. Lowenstein and co-workers have recently presented data supporting a role for pp60^{C-SRC} in gap junctions [19], which have also been shown experimentally to be playing an important role in determination and differentiation events in *Hydra* [86,116]. In addition, Loewenstein and co-workers have shown that the growth inhibition exerted upon transformed cells by normal cells correlates with the degree of junctional communication between these cells [200]. Therefore, the data available to date clearly supports pp60^{C-SRC} playing more of a role in differentiation and/or the maintenance of the differentiated state than in cellular proliferation. It would also appear that some form of cell-cell communication is central to this process, be it for example synaptical transmission or gap-junctional contact. This is also consistent with the sub-cellular localization of pp60^{C-SRC} underneath the plasma membrane. As the neoplastically transformed cell is characterized by an aberrant response to extracellular growth-regulating signals, it can be seen that the oncogenic potential of the c-src gene may merely be a pathologic extension of its normal function.

V. SUMMARY:

The isolation of several members of the *src* gene family from the teleost fish *Xiphophorus maculatus* has allowed an examination of three areas relating to these genes: 1) the sequence and gene topology relationship of these genes in a lower vertebrate to the homologous genes in invertebrates, higher vertebrates and mammals; 2) the relationship of the *Xiphophorus c-src* gene to the locus (*mdl*) responsible for the melanoma formation in the platyfish/swordtail hybrid fish; and 3) certain aspects of the evolutionary history of the *src* gene family, including the calculation of the approximate time of the *src/yes* gene duplication. The results can be summarized as follows:

1) The *Xiphophorus* genome, analogous to the situation in higher vertebrates and mammals, contains a family of *src*-related sequences. Whether the fish contain the full complement of characterized *src* family members of the higher vertebrates is unclear. Hybridization analyses suggested that several of the *Xiphophorus src*-related clones correspond to known gene family members - eg. 19-4 (*c-src*), 22-1 (*c-yes*) and 22-2 (*c-lck*). Sequence data from clones 19-4 and 22-1 (G. Hannig, unpublished) confirm their derivation from the fish *c-src* and *c-yes* genes respectively. The sequence of clone 19-4 (*Xsrc*) leaves little doubt that the corresponding cellular locus encodes a tyrosine-specific protein kinase. The pattern of expression of the *Xsrc* gene [197, 246] is consistent with the level and pattern of expression of a previously characterized fish protein with tyrosine kinase activity and reactive with rabbit antisera specific to the pp60^{v-src} protein [23]. The comparison of the genomic sequence of *Xsrc* to genomic sequence from the chicken and human *c-src* genes, as well as the isolation of a *Xiphophorus c-src* cDNA clone, allowed the determination of the exon/intron configuration of the fish gene. Based upon the available genomic sequence, the *Xsrc* gene has the identical exon/intron configuration as the genes of the higher vertebrate *src* gene family, which differs from that found in the *Drosophila c-src* sequences.

2) The *Xiphophorus maculatus c-src* gene does not map to the X chromosome. Therefore the melanoma inducing locus (*mdl*) and *c-src* are not the same gene, and are not physically linked to one another. This suggests some form of interaction, direct or indirect, between the *X. maculatus mdl* locus and the *X. helleri c-src* gene in the tumor-bearing fish. The role that the *c-src* gene may or may not be playing in the induction of the melanoma is still unclear. However, there are indications that certain events involving the *c-src* gene (eg. demethylation and/or amplification) may be playing a role in the progression of the tumor to full malignancy. Further analyses of the *c-src* gene in the benign and malignant melanomas are required to answer this question.

3) Comparison of the fish *c-src* and *c-yes* sequences with the human *c-src* and *c-yes* sequences allowed a calculation of the evolutionary divergence rate for both genes. These rates are very similar for both genes, and are extremely low (i.e. the *c-src* and *c-yes* proteins are very highly conserved over evolutionary time - as high as some histone proteins). This very high evolutionary conservation is found for protooncogenes in general. The DNA divergence rates support a constant evolutionary molecular clock for the *c-src* gene (with the exception of the

mouse/rodent lineage). The amino acid divergence rates suggest a slow-down in divergence rate for the pp60^{c-src} protein over evolutionary time. The comparison of *src* family member sequences and exon/intron configuration supports a single progenitor sequence for the vertebrate *src* gene family, as well as an early split of this gene family into two major branches. The time of the *src/yes* gene duplication could be approximated (550 ± 50 Myr ago), coinciding with the emergence of the first proto-chordates. This calculation is consistent with results in other systems suggesting a burst of gene duplication 500-600 Myr ago, with the blossoming of the fossil record during this time period, and with proposed chordate evolution scenarios involving several genome duplication events. The further characterization of the *src*-related sequences present in the genomes from fish, *Xenopus*, *C. elegans*, *Hydra*, and sponge should lead to an expansion of our knowledge concerning the evolution of the *src* gene and the *src* gene family, as well as giving rise to simpler model developmental systems with which to study the normal function of these genes.

VI. ZUSAMMENFASSUNG

Die Isolierung verschiedener Vertreter der *src* -Genfamilie des Knochenfisches *Xiphophorus maculatus* erlaubte Untersuchungen: 1) über die Verwandtschaftsbeziehungen der Sequenz und der Gen-Topologie dieser Gene eines niederen Vertebraten zu den homologen Genen von Invertebraten und höheren Vertebraten; 2) über die Beziehung des *c-src* Gens von *Xiphophorus* zu dem Gen-Locus *mdl*, der für die Melanombildung in Hybriden von Platyfish und Schwerträger verantwortlich ist; und 3), über verschiedene Aspekte der Evolution der *src* -Genfamilie, einschließlich der Berechnung des Zeitpunktes der *src* /*yes* Genduplikation. Die Resultate sind zusammengefasst folgende:

1) Das Genom von *Xiphophorus* enthält eine Genfamilie von *src* -verwandten Sequenzen, analog zu der Situation bei höheren Vertebraten. Ob der Fisch sämtliche Mitglieder der bisher charakterisierten *src* -Genfamilie höherer Vertebraten enthält, ist unklar. Hybridisierungsanalysen deuten jedoch darauf hin, daß verschiedene der *src* -verwandten isolierten Klone von *Xiphophorus* homolog sind zu verschiedenen Mitgliedern der *src* -Genfamilie höherer Vertebraten, so 19-4 (*c-src*), 22-1 (*c-yes*) und 22-2 (*c-lck*). Sequenz-Daten der Klone 19-4 und 22-1 (G. Hannig, unveröffentlicht) bestätigen, daß diese Klone vom Fisch *c-src* bzw. vom Fisch *c-yes* stammen. Die Sequenz von Klon 19-4 (*Xsrc*) codiert zweifelsfrei für eine Tyrosin-spezifische Proteinkinase. Die Expression des *Xsrc* -Gens [197,246] ist identisch mit dem Expressionsmuster eines früher charakterisierten Fischproteins mit Tyrosinkinase-Aktivität, das spezifisch mit Kaninchen-Antiseren gegen pp60^{v-src} reagierte [23]. Der Vergleich der genomischen Sequenz von *Xsrc* mit den genomischen Sequenzen von *c-src* von Mensch und Huhn, wie auch die Isolierung einer *c-src* cDNA von *Xiphophorus*, erlaubte die Bestimmung des genauen Exon/Intron-Arrangements des Fischgens. Basierend auf den verfügbaren genomischen Sequenzen, zeigt sich für das *Xsrc* Gen eine identische Exon/Intron-Konfiguration wie bei den homologen Genen höherer Vertebraten, die von den *Drosophila c-src* Sequenzen abweicht.

2) Das *c-src* Gen von *Xiphophorus maculatus* (*Xsrc*) ist nicht auf dem X-Chromosom lokalisiert. Aus diesem Grunde sind der Melanom-induzierende Locus *mdl* und *Xsrc* nicht identisch und nicht miteinander gekoppelt. Dies läßt ein direktes oder indirektes Zusammenspiel zwischen dem *mdl*-Locus von *Xiphophorus maculatus* und dem *c-src* Gen von *X. helleri* in Tumor-tragenden Fischen vermuten. Die Rolle, die das *c-src* Gen für die Induktion von Melanomen spielt, bleibt weiterhin unklar. Trotzdem gibt es Hinweise darauf, daß bestimmte Ereignisse, wie eine Demethylierung und/oder eine Amplifizierung des *c-src* Gens, bei der Progression des Tumors zu voller Malignität einer Rolle spielen. Weitere Analysen des *c-src* Gens in benignen und malignen Melanomen sind für eine abschließende Beantwortung dieser Frage erforderlich.

3) Ein Vergleich der *c-src* und *c-yes* Sequenzen von Fisch und Mensch erlaubte eine Berechnung der evolutionären Divergenzrate für beide Gene. Diese Werte sind für beide Gene sehr ähnlich und außerordentlich niedrig. Das heißt, daß die *c-src* und *c-yes* Proteine über die entsprechende Evolutionszeit hinweg so hochkonserviert erscheinen wie einige Histon-Proteine. Diese sehr hohe Konservierung wird jedoch für Protoonkogene ganz allgemein

gefunden. Die DNA-Divergenzraten stützen die Annahme einer konstanten molekularen Uhr während der Evolution für das *c-src* Gen, mit der Ausnahme der Maus/Nagetier-Linie. Die Aminosäure-Divergenzraten lassen eine Verlangsamung der Divergenzraten während der Evolutionszeit für das pp60^{c-src} Protein vermuten. Der Vergleich von Mitgliedern der *src* -Genfamilie und ihren Exon/Intron-Arrangements stärkt die Annahme, daß es eine einzige Ursequenz für die gesamte *src* -Genfamilie der Vertebraten gegeben hat, und daß es eine frühe Aufspaltung dieser Genfamilie in zwei Hauptzweige gegeben hat. Die *src/yes* -Duplikation konnte näherungsweise auf 550 (± 50) Millionen Jahre zurückdatiert werden, auf einen Zeitpunkt als die ersten Protochordaten auftraten. Diese Berechnung steht in Übereinstimmung mit Daten aus anderen Systemen und paläontologischen Funden, die eine Vielzahl von Genduplikationen vor 500-600 Millionen Jahren nahelegen. Die weitere Charakterisierung von *src* -verwandten Sequenzen von Fisch, *Xenopus*, *C. elegans*, *Hydra* und Schwamm sollte eine Erweiterung unseres Wissens über die Evolution des *src* -Gens und der *src* -Genfamilie bringen, ebenso sollte die Beschäftigung mit Vertretern der Invertebraten ein einfacheres entwicklungsbiologisches Modellsystem aufzeigen, mit dem die normale Funktion dieser Gene studiert werden kann.

VII REFERENCES

1. Abelson, H. T. and Rabstein, L. S. "Lymphosarcoma: virus-induced thymic-independent disease in mice" *Cancer Res.* 30: 2213-2222 (1970a)
2. Adam, D., Wittbrodt, J., Telling, A. and Scharlt, M. "RFLP for an EGF-receptor related gene associated with the melanoma oncogene locus of *Xiphophorus maculatus*" *Nucl. Acids Res.* (in press) 1988
3. Adams, R. L. P. and Burdon, R. H. 'Molecular biology of DNA methylation' Springer-Verlag, N.Y. 1985 pp 115-162
4. Adams, R. L. P. and Burdon, R. H. 'Molecular Biology of DNA Methylation' Springer-Verlag, N. Y. 1985. pp 171-174
5. Ahnn, J., March, P. E., Takiff, H. E. and Inouye, M. "A GTP-binding protein of *Eschericia coli* has homology to yeast RAS proteins" *Proc. Natl. Acad. Sci. U.S.A* 83: 8849-8853 (1986)
6. Alema, S., Casalbone, P., Agostini, E. and Tato, F. "Differentiation of PC12 phaeochromocytoma cells induced by *v-src* oncogene" *Nature* 316: 557-559 (1985)
7. Altitalo, K. "Amplification of cellular oncogenes in cancer cells" *TIBS*, May, 1985. pp 194-197
8. Ames, B. N. and Hooper, K. "Does carcinogenic potency correlate with mutagenic potency in the Ames assay" *Nature* 274: 19-20 (1978)
9. Anders, A. and Anders, F. "Etiology of cancer as studied in the platyfish-swordtail system" *Biochim. Biophys. Acta* 516: 61-95 (1978)
10. Anders, A., Anders, F. and Klinke, K. "Regulation of gene expression in the Gordon-Kosswig melanoma system. I. The distribution of controlling genes in the genome of the Xiphophorin fish, *Platyopocilus maculatus* and *Platyopocilus variatus*" in: 'Genetics and Mutagenesis of Fish' Schroeder, H. (ed.), Springer, Berlin. 1973. pp 33-52
11. Anders, F. "Tumour formation in platyfish-swordtail hybrids as a problem of gene regulation" *Experientia* 23: 1-10 (1967)
12. Anders, F., Anders, A., Scharlt, M., Gronau, T., Lüke, W., Schmidt, C-R. and Barnekow, A. "Oncogenes in development, neoplasia, and evolution" in: 'New Experimental Modalities in the control of Neoplasia' P. Chandra (ed.) Plenum Press, N. Y. 1986 pp 15-40
13. Anders, F., Gronau, T., Scharlt, M., Barnekow, A., Jaenel-Dess, G. and Anders, A. "Cellular oncogenes as ubiquitous genomic constituents in the animal kingdom and as fundamentals in melanoma formation" in: 'Proceedings of the First International Conference on Skin Melanoma' N. Cascinelli (ed.)
14. Anders, F., Scharlt, M., Barnekow, A. and Anders, A. "Xiphophorus as an in vivo model for studies on normal and defective control of oncogenes" *Adv. Cancer Res.* 42: 191-275 (1984)
15. Anders, F. "Onkogene in Evolution, Entwicklung und Tumorbildung" in: 'Zellenergetic und Zelldifferenzierung, Ergebnisbericht des Sonderforschungsbereichs 103, 1983-1985. Philipps Universität, Marburg. pp 172-186
16. Anderson, S. K., Gibbs, C. P., Tanaka, A., Kung, H-J. and Fujita, D. J. "Human cellular src

- gene: nucleotide sequence and derived amino acid sequence of the region coding for the carboxy-terminal two-thirds of pp60c-src" *Molec. Cell. Biol.* 5: 1122-1129 (1985)
17. Armelin, H. A., Armelin, M.C.S., Kelly, K., Stewart, T., Leder, P., Cochran, B.H. and Stiles, C.D. "Functional role for c-myc in mitogenic response to platelet-derived growth factor" *Nature* 310: 655-660 (1984)
 18. Atz, J. W. "Effects of hybridization on pigmentation in fishes of the genus *Xiphophorus*" *Zoologica* 47: 153-190 (1962)
 19. Azarnia, R., Reddy, S., Kmiecik, T. E., Shalloway, D. and Loewenstein, W. R. "The cellular src gene product regulates junctional cell-to-cell communication" *Science* 239: 398-401 (1988)
 - 19a. Bairoch, A. and Claverie, J. M. "Sequence patterns in protein kinases" *Nature* 331: 22 (1988)
 20. Bargmann, C. I., Hung, M.C. and Weinberg, R. A. "Multiple independent activations of the neu oncogene by a point mutation altering the transmembrane domain of p185" *Cell* 45: 649-657 (1986)
 21. Barker, W. C. and Dayhoff, M. O. "Viral src gene products are related to the catalytic chain of mammalian cAMP-dependent protein kinase" *Proc. Natl. Acad. Sci. U.S.A* 79: 2836-2839 (1982)
 - 21a. Barnekow, A. and Gessler, M. "Activation of the pp60c-src kinase during differentiation of monomyelocytic cells in vitro" *EMBO J.* 5: 701-705 (1986)
 22. Barnekow, A. and Scharf, M. "Cellular src gene product detected in the freshwater sponge *Spongilla lacustris*" *Molec. Cell. Biol.* 4: 1179-1181 (1984)
 23. Barnekow, A., Scharf, M., Anders, F. and Bauer, H. "Identification of a fish protein associated with a kinase activity and related to the Rous sarcoma virus transforming protein" *Cancer Res.* 42: 2429-2433 (1982)
 24. Baylin, S. B. "The multiple endocrine neoplasia syndromes: implications for the study of inherited tumors" *Sem. Oncol.* 5: 35-45 (1978)
 - 24a. Beltz, G. A., Jacobs, K. A., Eickbush, T. H., Cherbash, P. T. and Kafatos, F. C. "Isolation of multigene families and determination of homologies by filter hybridization methods" *Meth. Enzymol.* 100: 266-285 (1983)
 - 24b. Bender, W. and Peifer, M. "Oncogenes take wing" *Cell* 50: 519-520 (1987)
 25. Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F. "The mosaic genome of warm-blooded vertebrates" *Science* 228: 953-958 (1985)
 26. Bernardi, G. and Bernardi, G. "Compositional constraints and genome evolution" *J. Mol. Evol.* 24: 1-11 (1986)
 27. Besmer, P., Murphy, J. E., George, P. C., Oiu, F., Bergold, P. J., Lederman, L., Snyder, H. W. Jr., Brodeur, D., Zuckerman, E. E. and Hardy, W. D. "A new acute transforming feline retrovirus and relationship of its oncogene v-kit with the protein kinase gene family" *Nature* 320: 415-421 (1986)
 - 27a. Biggin, M. D., Gibson, T. J. and Hong, G. F. "Buffer gradient gels and 35S label as an aid to rapid DNA sequence determination" *Proc. Natl. Acad. Sci. U.S.A.* 80: 3963-3965

(1983)

26. Bird, A. P. "CpG-rich Islands and the function of DNA methylation" *Nature* 321: 209-213 (1986)
29. Bird, A. P. and Taggart, M. H. "Variable patterns of total DNA and rDNA methylation in animals" *Nucl. Acids Res.* 8: 1485-1497 (1980)
30. Bird, A. P., Taggart, M., Frommer, M., Miller, O. J. and Macleod, D. "A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA" *Cell* 40: 91-99 (1985)
31. Bird, A. P. "CpG Islands as gene markers in the vertebrate nucleus" *Trends Genet.* 3: 342-347 (1987)
- 31a. Birnboim, H. C. "A rapid alkaline extraction method for the isolation of plasmid DNA" *Meth. Enzymol.* 100: 243-255 (1983)
32. Bishop, J. M. "Viral oncogenes" *Cell* 42: 23-38 (1985)
33. Bishop, J. M. "Cellular oncogenes and retroviruses" *Ann. Rev. Biochem.* 52: 301-354 (1983)
34. Bishop, J. M. "The molecular genetics of cancer" *Science* 235: 305-311 (1987)
- 34a. Bishop, J.M. and Varmus, H.E. "Functions and origins of retroviral transforming genes" in 'Molecular Biology of Tumor Viruses: RNA tumor viruses' Weiss, R. et al (eds.) Cold Spring Harbor Laboratory, 1982. pp999
35. Blake, C. C. F. "Exons and the evolution of proteins" *Intl. Rev. Cytol.* 93: 149-185 (1985)
36. Blattner, W. A. "Familial cancer: an opportunity to study mechanisms of neoplastic transformation" in: 'Oncodevelopmental Markers: Biologic, Diagnostic and Monitoring Aspects' Academic Press, N.Y. 1983. pp 189-204
- 36a. Blin, N. and Stafford, D. W. "A general method for isolation of high molecular weight DNA from eukaryotes" *Nucl. Acids Res.* 3: 2303-2308 (1976)
37. Bode, H., Dunne, J., Heimfeld, S., Huang, L., Javois, L., Koizumi, O., Westerfield, J. and Yaross, M. "Transdifferentiation occurs continuously in adult hydra" *Curr. Top. Devel. Biol.* 20: 257-280 (1986)
38. Bodmer, W. F. "Cancer genetics" *Cancer Surveys* 1: 1-15 (1982)
39. Bourne, H. R. and Sullivan, K. A. "Mammalian G proteins: models for ras proteins in transmembrane signalling?" *Cancer Surveys* 5: 257-274 (1986)
40. Boveri, T. 'Zür Frage der Entstehung maligner Tumoren' Verlag von Gustav Fischer, Jena. 1914
41. Breider, H. "Über Melanosarkome, Melaninbildung und homologe Zellmechanismen" *Strahlentherapie* 88: 619-639 (1952)
42. Britten, R. J. "Rates of DNA sequence evolution differ between taxonomic groups" *Science* 231: 1393-1398 (1986)
43. Burnett, A. L., Lowell, R. and Cyrlin, M. N. "Regeneration of a complete Hydra from a single, differentiated somatic cell type" in: 'Biology of Hydra' A. L. Burnett (ed.) Academic Press, N

Y. 1973. pp 255-270

44. Burt, R. W. et al "Dominant inheritance of adenomatous colonic polyps and colorectal cancer" *New Engl. J. Med.* 312: 1540-1544 (1985)
45. Busslinger, M., Hurst, J. and Flavell, R. A. "DNA methylation and the regulation of globin gene expression" *Cell* 34: 197-206 (1983)
46. Cairns, J. "Mutation, selection and the natural history of cancer" *Nature* 255: 197-200 (1975)
47. Cairns, J. "The origin of human cancers" *Nature* 289: 353-357 (1981)
48. Cavanee, W. K., Dryja, T. P., Phillips, R. A., Benedict, W. F., Godbout, R., Gallie, B. L., Murphree, A. L., Strong, L. C. and White, R. L. "Expression of recessive alleles by chromosomal mechanisms in retinoblastoma" *Nature* 305: 779-784 (1983)
49. Cedar, H. "DNA methylation and gene activity" *Cell* 53: 3-4 (1988)
50. Cedar, H. "DNA methylation and gene expression" in: 'DNA methylation: biochemistry and biological significance' Razin, A., Cedar, H. and Riggs, A. D. (eds.) Springer-Verlag, N.Y. 1984. pp 147-164
51. Chandler, L. A., DeClerck, Y. A., Bogenmann, E. and Jones, P. A. "Patterns of DNA methylation and gene expression in human tumor cell lines" *Cancer Res.* 46: 2944-2949 (1986)
52. Chang, C. and Meyerowitz, E. M. "Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene" *Proc. Natl. Acad. Sci. U.S.A* 83: 1408 (1986)
53. Chatterjee, K., Kollinger, G., Schmidt, C-R., Anders, A. and Anders, F. "Cytogenetics of neoplasia of Xiphophorus" *Cancer Genet. Cytogenet.* 3: 195-209 (1981)
- 53a. Chen, E. Y. and Seeburg, P. H. "Supercoil sequencing: a fast and simple method for sequencing plasmid DNA" *DNA* 4: 165-170 (1985)
54. Cooper, J. A. "Transforming mutations in protein-tyrosine kinase genes" *Bioessays* 4: 9-15 (1986)
55. Cooper, J. A., Gould, K. L., Cartwright, C. A. and Hunter, T. "Tyr527 is phosphorylated in pp60c-src: implications for regulation" *Science* 231: 1431-1434 (1986)
56. Cotton, P. C. and Brugge, J. S. "Neural Tissues express high levels of the cellular src gene product pp60c-src" *Molec. Cell. Biol.* 3: 1157-1162 (1983)
57. Cross, F. R., Garber, E. A., Pellman, D. and Hanafusa, H. "A short sequence in the pp60^{src} N-terminus is required for pp60^{src} myristylation and membrane association and for cell transformation" *Molec. Cell. Biol.* 4: 1834-1842 (1984)
- 57a. Crouse, G. F., Frischauf, A. and Lehrach, H. "An integrated and simplified approach to cloning into plasmids and single-stranded phages" *Meth. Enzymol.* 101: 78-89 (1983)
58. Czernilofsky, A. P., Levinson, A. D., Varmus, H. E., Bishop, J. M., Tischler, E. and Goodman, H. "Corrections to the nucleotide sequence of the src gene of Rous sarcoma virus" *Nature* 301: 736-738 (1983)
59. Davie, J. R., Delcuve, G. P., Nickel, B. E., Moirier, R. and Bailey, G. "Reduced levels of histones H1^o and H1b, and unaltered content of methylated DNA in rainbow trout hepatocellular carcinoma chromatin" *Cancer Res.* 47: 5407-5410 (1987)

60. Dawe, C. J. "Phylogeny and oncogeny" in: 'Neoplasms and Related Disorders of Invertebrate and Lower Vertebrate Animals' Natl. Cancer Inst. Monograph 31, 1969. pp 1-40
61. Dayhoff, M. O. 'Atlas of Protein Sequence and Structure' Vol. 5 (Suppl. 3) Natl. Biomed. Res. Found'n. 1978.
62. DeFeo-Jones, D., Scolnick, E., Koller, R. and Dhar, R. "ras -related gene sequences identified and isolated from *Saccharomyces cerevisiae*" Nature 306: 707-709 (1983)
63. de Gouvea, H. Ann. Ocul. 143: 32-34 (1910)
64. Delli Bovi, P., Curatola, A. M., Kern, F. G., Greco, A., Iltmann, M. and Basilico, C. "An oncogene isolated by transfection of Kaposi's sarcoma DNA encodes a growth factor that is a member of the FGF family" Cell 50: 729-737 (1987)
65. DeLorbe, W. J., Luciw, P. A., Goodman, H. M., Varmus, H. E. and Bishop, J. M. "Molecular cloning and characterization of avian sarcoma virus circular DNA molecules" J. Virol. 36: 50-61 (1980)
66. Der, C. J., Krontiris, T. G. and Cooper, G. M. "Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses" Proc. Natl. Acad. Sci. U.S.A 79: 3637-3640 (1982)
67. Devare, S. G., Reddy, E. P., Law, J. D., Robbins, K. C. and Aaronson, S. A. "Nucleotide sequence of the simian sarcoma virus genome: demonstration that its acquired cellular sequences encode the transforming gene product p28sis" Proc. Natl. Acad. Sci. U.S.A 80: 731-735 (1983)
- 67a. Devereux, Haeberli and Smithies "A comprehensive set of sequence analysis programs for the VAX" Nucl. Acids Res. 12: 387-395 (1984)
68. Dials, E. S., Cheah, M. S. C., Rowitch, D. and Hoffman, R. M. "Extent of DNA methylation in human tumor cells" J. Natl. Cancer Inst. 71: 755-764 (1983)
69. DiBerardino, M. A., Mizell, M., Hoffner, N. J. and Friesendorf, D. G. "Frog larvae cloned from nuclei of pronephric adenocarcinoma" Differentiation 23: 213-217 (1983)
70. Dickson, C. and Peters, G. "Potential oncogene product related to growth factors" Nature 326: 833 (1987)
71. Diehl, F. A. "The developmental significance of interstitial cells during regeneration and budding" in: 'Biology of Hydra' A. L. Burnett (ed.) Academic Press, N. Y. 1973. pp 109-143
72. Doerfler, W. "DNA methylation and gene activity" Ann. Rev. Biochem. 52: 93-124 (1983)
73. Donner, L., Fedele, L. A., Garon, C. F., Anderson, S. J. and Sherr, C. J. "McDonough feline sarcoma virus: characterization of the molecularly cloned provirus and its feline oncogene (v-fms)" J. Virol. 41: 489-500 (1982)
74. Doolittle, R. F. "The genealogy of some recently evolved vertebrate proteins" TIBS, June 1985. pp 233-237
75. Duesberg, P. H. and Vogt, P. K. "Differences between the ribonucleic acids of transforming and nontransforming avian tumor viruses" PNAS 67: 1673-1680 (1970)
76. Duesberg, P. H. "Retroviral transforming genes in normal cells?" Nature 304: 219-226

(1983)

77. Dunnick, W., Baumgartner, J. Fradkin, L., Schultz, C. and Szurek, P. "Methylation of plasmacytoma *c-myc* genes" *Gene* 39: 287-292 (1985)
78. Dyson, P. J., Quade, K. and Wyke, J. A. "Expression of the ASV *src* gene in hybrids between normal and virally transformed cells: Specific suppression occurs in some hybrids but not in others" *Cell* 30: 491-498 (1982)
79. Eisenman, R. N. and Thompson, C. B. "Oncogenes with potential nuclear function: *myc*, *myb* and *fos*" *Cancer Surveys* 5: 309-327 (1986)
80. Evans, E. P., Burtenshaw, M. D., Brown, B. B., Hennion, R. and Harris, H. "The analysis of malignancy by cell fusion. IX. Re-examination and clarification of the cytogenetic problem" *J. Cell Sci.* 56: 113-130 (1982)
81. Farber, E. and Cameron, R. "The sequential analysis of cancer development" *Advances in Cancer Research* 31: 125-226 (1980)
82. Farber, E. "The multistep nature of cancer development" *Cancer Res.* 44: 4217-4223 (1984)
83. Feinberg, A. and Vogelstein, B. "Hypomethylation of *ras* oncogenes in primary human cancers" *Biochem. Biophys. Res. Comm.* 111: 47-54 (1983a)
84. Feinberg, A. P. and Vogelstein, B. "Hypomethylation distinguishes genes of some human cancers from their normal counterparts" *Nature* 301: 89-92 (1983)
- 83a. Feinberg, A. P. and Vogelstein, B. "A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity" *Anal. Biochem.* 132: 6-13 (1983)
85. Feramisco, J. R., Kamata, T., Gross, M., Rasenberg, M. and Sweet, R. W. "Microinjection of the oncogene form of the human *H-ras* (T24) protein results in rapid proliferation of quiescent cells" *Cell* 38: 109-117 (1984)
86. Fraser, S. E., Green, C. R., Bode, H. and Gilula, N. B. "Selective disruption of gap junctional communication interferes with a patterning process in hydra" *Science* 237: 49-55 (1987)
87. Fried, M. "Cell transformation ability of a temperature-sensitive mutant of polyoma virus" *Proc. Natl. Acad. Sci. U.S.A* 53: 486-491 (1965)
88. Friedman, B. and Gordon, M. "Chromosome numbers in xiphophore fish" *Am. Nat.* 58: 446-455 (1934)
89. Friend, S. H., Bernards, R., Rogelj, S., Weinberg, R. A., Rapaport, J. M., Albert, D. M. and Dryja, T. P. "A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma" *Nature* 323: 643-646 (1986)
- 89a. Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N. "Lambda replacement vectors carrying polylinker sequences" *J. Mol. Biol.* 170: 827-842 (1983)
90. Fults, D. W., Towle, A. C., Lauder, J. M. and Maness, P. F. "*pp60^{c-src}* in the developing cerebellum" *Molec. Cell. Biol.* 5: 27-32 (1985)
- 90a. Gallie, B.L., Ellsworth, R.M., Abramson, D.H. and Phillips, R.A. "Retinoma: spontaneous regression of retinoblastoma or benign manifestation of the mutation?" *Br. J. Cancer* 45: 513-521 (1982)

91. Gallwitz, D., Donath, C. and Sander, C. "A yeast gene encoding a protein homologous to the human c-ha/bas proto-oncogene product" *Nature* 306: 704-707 (1983)
92. Galtsoff, P. S. "Regeneration after dissociation (an experimental study on sponges). I. Behavior of dissociated cells of *Microclona prolifera* under normal and altered conditions" *J. Exp. Zool.* 42: 183-221 (1925)
93. Gama-Sosa, M. A., Slagel, V. A., Trewyn, R. W., Oxenhandler, R., Kuo, K., Gehrke, C. W. and Ehrlich, M. "The 5-methylcytosine content of DNA from human tumors" *Nucl. Acids Res.* 11: 6883-6894 (1983)
- 93a. Garoff, H. and Ansorge, W. "Improvements of DNA sequencing gels" *Anal. Biochem.* 115: 450-457 (1981)
94. Gateff, E. "Malignant neoplasms of genetic origin in *Drosophila melanogaster*" *Science* 200: 1448-1459 (1978)
95. Gerschenson, M., Graves, K., Carson, S. D., Wells, R. S. and Pierce, G. B. "Regulation of melanoma by the embryonic skin" *Proc. Natl. Acad. Sci. U.S.A* 83: 7307-7310 (1986)
96. Gilbert, W. "The exon theory of genes" *Cold Spring Harbor Symp. Quant. Biol.* 52: 901-905 (1987)
97. Gilbert, W., Marchionni, M. and McKnight, G. "On the antiquity of introns" *Cell* 46: 151-154 (1986)
98. Goddard, J. M., Welland, J. J. and Capecchi, M. R. "Isolation and characterization of *Caenorhabditis elegans* DNA sequences homologous to the v-abl oncogene" *Proc. Natl. Acad. Sci. U.S.A* 83: 2172-2176 (1986)
99. Goelz, S. E., Vogelstein, B., Hamilton, S. R. and Feinberg, A. P. "Hypomethylation of DNA from benign and malignant human colon neoplasms" *Science* 228: 187-190 (1985)
100. Goff, S. P. and Baltimore, D. in: *Advances in Viral Oncology Vol. 1* G. Klein (ed.) Raven Press, N.Y. 1982. pp 127-139
101. Goff, S. P., Gilboa, E., Witte, O. N. and Baltimore, D. "Structure of the Abelson murine leukemia virus genome and the homologous cellular gene: studies with cloned viral DNA" *Cell* 22: 777-785 (1980)
102. Gojobori, T. "Evolutionary features of oncogenes" in: *Population Genetics and Molecular Evolution* T. Ohta and K. Aoki (eds.) Springer-Verlag, Berlin. 1985. pp 353-368
103. Gojobori, T. and Yokoyama, S. "Molecular evolutionary rates of oncogenes" *J. Mol. Evol.* 26: 148-156 (1987)
- 103a. Gomer, R., Datta, S. and Firtel, R. *Focus* 7: 1,6 (1985)
104. Gootwine, E., Webb, C. G. and Sachs, L. "Participation of myeloid leukaemic cells injected into embryos in haematopoietic differentiation in adult mice" *Nature* 299: 63-65 (1982)
105. Gordon, M. "The genetics of a viviparous top-minnow *Platypharodon* ; The inheritance of two kinds of melanophores" *Genetics* 12: 253-283 (1927)
106. Gordon, M. "Hereditary basis of melanosis in hybrid fishes" *Am J. Cancer* 15: 1495-1523 (1931)
107. Gordon, M. "The hereditary basis for melanosis in hybrids of Mexican killifishes" *Proc*

108. Gordon, M. "The development of melanomas from macromelanophores in different body regions under genetic control" *Anat. Rec.* 89: 555 (1944)
109. Gordon, M. "Genetic and correlated studies of normal and atypical pigment cell growth" 10th Symp. Soc. for the Study of Development and Growth, Northhampton, Mass. 1951. pp 153-219 (1951)
110. Gordon, M. "The variable expressivity of a pigment cell gene from zero effect to melanotic tumor induction" *Cancer Res.* 11: 676-686 (1951)
111. Gordon, M. "An intricate genetic system that controls nine pigment cell patterns in the platyfish" *Zoologica* 41: 153-162 (1956)
112. Gordon, M. "The origin, development, but incomplete differentiation of pigment cells in melanomas" *Proc. Am. Assoc. Cancer Res.* 2: 207 (1957)
113. Gordon, M. "A genetic concept for the origin of melanomas" *Ann. N. Y. Acad. Sci.* 71: 1213-1222 (1958)
- 113a. Gould, K.L. and Hunter, T. "Platelet-derived growth factor induces multisite phosphorylation of pp60c-src and increases its protein-tyrosine kinase activity" *Molec. Cell. Biol.* 8: 3345-3356 (1988)
114. Goustin, A. S., Leof, E. B., Shipley, G. D. and Moses, H. L. "Growth factors and cancer" *Cancer Res.* 46: 1015-1029 (1986)
- 114a. Graf, T., Beug, H. and Hayman, M.J. "Target cell specificity of defective avian leukemia viruses: hematopoietic target cells for a given virus type can be infected but not transformed by strains of a different type" *Proc. Natl. Acad. Sci. U.S.A.* 77: 389-393 (1980)
115. Grand, C. G., Gordon, M. and Cameron, G. "Neoplasm studies. VIII. Cell types in tissue culture of fish melanotic tumors compared with mammalian melanomas" *Cancer Res.* 1: 660-666 (1941)
116. Green, C. R. "Evidence mounts for the role of gap junctions during development" *BioEssays* 8: 7-10 (1988)
117. Greene, M. H. and Fraumeni, J. F. "The hereditary variant of malignant melanoma" in: 'Human Malignant Melanoma' Clark, W.H., Goldman, L.I. and Mastrangelo, M.J. (eds.) Grune et Stratton, New York. 1979. pp 139-166
118. Gregory, R. J., Kammermeyer, K. L., Vincent, W. S., Wadsworth, S. G. "Primary sequence and developmental expression of a novel *Drosophila melanogaster* src gene" *Mol. Cell. Biol.* 7: 2119-2127 (1987)
119. Groudine, M. and Weintraub, H. "Activation of globin genes during chicken development" *Cell* 24: 393-401 (1981)
120. Hafen, E., Basler, K., Edstroem, J-E. and Rubin, G. M. "Sevenless, a cell-specific homeotic gene of *Drosophila*, encodes a putative transmembrane receptor with a tyrosine kinase domain" *Science* 236: 55-63 (1987)
121. Haller, J., Cote, S., Brönner, G. and Jäckle, H. "Dorsal and neural expression of a tyrosine kinase-related *Drosophila* gene during embryonic development" *Genes Dev.* 1: 862-867 (1987)

122. Hanahan, D. "Heritable formation of pancreatic β -cell tumors in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes" *Nature* 315: 115-122 (1985)
123. Hanahan, D. "Oncogenesis in transgenic mice" In: 'Oncogenes and Growth Control' P. Kahn and T. Graf (eds.) Springer-Verlag, Berlin. 1986. pp 349-363
- 123a. Hanahan, D. "Techniques for transformation of *E. coli*" In: 'DNA Cloning' Vol. 2. D. Glover (ed.) Academic Press, N.Y. 1986. pp 109-135
124. Hanawalt, P. C. and Sarasin, A. "Cancer-prone hereditary diseases with DNA processing abnormalities" *Trends in Genetics*, May 1986. pp 124-129
- 124a. Hanks, S.K., Quinn, A.M. and Hunter, T. "The protein kinase family: conserved features and deduced phylogeny of the catalytic domains" *Science* 241: 42-52 (1988)
125. Hansen, M. F., Koufos, A., Gallie, B. L., Phillips, R. A., Fodstad, O., Brogger, A., Gedde-Dahl, T. and Cavane, W. K. "Osteosarcoma and retinoblastoma: a shared chromosomal mechanism revealing recessive predisposition" *Proc. Natl. Acad. Sci. U.S.A.* 82: 6216-6220 (1985)
126. Harland, R. M. "Inheritance of DNA methylation in microinjected eggs of *Xenopus laevis*" *Proc. Natl. Acad. Sci. U.S.A.* 79: 2323-2327 (1982)
127. Harris, M. "Induction of thymidine kinase in enzyme-deficient Chinese hamster cells" *Cell* 29: 483-492 (1982)
128. Harris, H., Miller, O. J., Klein, G., Worst, P. and Tachibana, T. "Suppression of malignancy by cell fusion" *Nature* 223: 363-368 (1969)
129. Harris, H. "The genetic analysis of malignancy" *J. Cell Sci. Suppl.* 4: 431-444 (1986)
130. Harvey, J. J. "An unidentified virus which causes the rapid production of tumors in mice" *Nature* 204: 1104-1105 (1964)
131. Häussler, G. "Über Melanombildungen bei Bastarden von *Xiphophorus helleri* und *Platyopocilus maculatus* var. *rubra*" *Klin. Wchnschr.* 7:1561-1562 (1928)
- 131a. Helms, C., Graham, M. Y., Dutchik, J. E. and Olson, M. V. "A new method for purifying lambda DNA from phage lysates" *DNA* 4: 39-49 (1985)
132. Henkemeyer, M. J., Bennett, R. L., Gertler, F. B. and Hoffmann, F. M. "DNA sequence, structure, and tyrosine kinase activity of the *Drosophila melanogaster* Abelson proto-oncogene homolog" *Mol. Cell. Biol.* 8: 843-853 (1988)
133. Heston, W. E. "Genetics: animal tumors" In: 'Cancer: A Comprehensive Treatise' 2nd Edition. F. F. Becker (ed.) Plenum Press, N. Y. 1982 pp 47-71
134. Holliday, R. "A new theory of carcinogenesis" *Br. J. Cancer* 40: 513-522 (1979)
135. Holliday, R. "The inheritance of epigenetic defects" *Science* 238: 163-170 (1987)
136. Howell, N. and Sager, R. "Non-coordinate expression of SV40-induced transformation and tumorigenicity in mouse cell hybrids" *Somatic Cell Genet.* 5: 129-143 (1979)
137. Howell, N. and Sager, R. "Genetic analysis of tumorigenesis: VIII. Suppression of SV40 transformation in cell hybrids and cytoplasmic transferants" *Cytogenet. Cell Genet.* 31: 214-227 (1981)

138. Huang, C-C., Hammond, C. and Bishop, J. M. "Nucleotide sequence of *v-fps* in the PRCII strain of avian sarcoma virus" J. Virol. 50: 125-131 (1984)
139. Huebner, R. J. and Todaro, G. J. "Oncogenes of RNA tumor viruses as determinants of cancer" Proc. Natl. Acad. Sci. U.S.A 64: 1087-1094 (1969)
140. Hughes, S. H., Payvar, F., Spector, D., Schinke, R.T., Robinson, H.L., Payne, G.S., Bishop, J.M. and Varmus, H.E. "Heterogeneity of genetic loci in chickens: analysis of endogenous viral and non-viral genes by cleavage of DNA with restriction enzymes" Cell 18: 347-359 (1979)
141. Humm, D. G. and Young, R. S. "The embryological origin of pigment cells in platyfish-swordtail hybrids" Zoologica 41: 1-10 (1956)
142. Hunter, T. "A tail of two src's: mutatis mutandis" Cell 49: 1-4 (1987)
143. Hunter, T. "A thousand and one protein kinases" Cell 50: 823-829 (1987)
144. Hunter, T and Cooper, J. A. "Protein-tyrosine kinases" Ann. Rev. Biochem. 54: 897-930 (1985)
145. Iba, H., Takeya, T., Cross, F. R., Hanafusa, T and Hanafusa H. "Rous sarcoma variants that carry the cellular *src* gene instead of the viral *src* gene cannot transform chicken embryo fibroblasts" Proc. Natl. Acad. Sci. U.S.A 81: 4424-4428 (1984)
146. Jones, P. A. "DNA methylation and cancer" Cancer Res. 46: 461-466 (1986)
147. Jones, W. C. "Is there a nervous system in sponges?" Biol. Rev. 37: 1-50 (1962)
148. Jukes, T. H. "Silent nucleotide substitutions and the molecular evolutionary clock" Science 21: 973-978 (1980)
149. Jukes, T. H. and Holmquist, R. "Evolutionary clock: nonconstancy of rate in different species" Science 177: 530-532 (1972)
150. Kaiser, H. E. "Animal neoplasms - a systematic review" in: 'Neoplasms - Comparative Pathology of Growth in Animals, Plants and Man' H. E. Kaiser (ed.) Williams & Wilkins, London. 1981. pp747-812
151. Kallman, K. D. "The platyfish, *Xiphophorus maculatus*" in: 'Handbook of Genetics Vol. 4: Vertebrates of Genetic Interest' R. C. King (ed.) Plenum Press, N. Y. 1975. pp 81-132
152. Kallman, K. D. and Schreibman, M. P. "The origin and possible genetic control of new, stable pigment patterns in the poeciliid fish *Xiphophorus maculatus*" J. Exp. Zool. 176: 147-168 (1971)
153. Katamine, S., Notario, V., Rao, C. D., Miki, T., Cheah, M. S. C., Tronick, S. R. and Robbins, K. C. "Primary structure of the human *lgr* proto-oncogene product p55^{c-*lgr*}" Mol. Cell. Biol. 8: 259-266 (1988)
154. Kataoka, T., Powers, S., Cameron, S., Fasano, O., Goldfarb, M., Broach, J. and Wigler, M. "Functional homology of mammalian and yeast RAS genes" Cell 40: 19-26(1985)
155. Kawai, S., Yoshida, M., Segawa, K., Sugiyama, H., Ishizaki, R. and Toyoshima, K. "Characterization of Y73, an avian sarcoma virus: a unique transforming gene and its product, a phosphopolypeptide with protein kinase activity" Proc. Natl. Acad. Sci. U.S.A 77: 6199-6203 (1980)

156. Kawakami, T., Pennington, C. Y. and Robbins, K. C. "Isolation and oncogenic potential of a novel human *src*-like gene" *Molec. Cell. Biol.* 6: 4195-4201 (1986)
157. Kitamura, N., Kitamura, A., Toyoshima, K., Hirayama, Y. and Yoshida, M. "Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus" *Nature* 297: 205-208 (1982)
- 157a. Klein, G. and Klein, A. "Evolution of tumors and the impact of molecular oncology" *Nature* 315: 190-195 (1985)
158. Klein, G. "The approaching era of the tumor suppressor genes" *Science* 238: 1539-1545 (1987)
159. Klemsz, M. J., McKercher, S. R. and Maki, R. A. "Nucleotide sequence of the mouse *hck* gene" *Nucl. Acids Res.* 15: 9600 (1987)
160. Knudson, A. G. "Mutation and cancer: statistical study of retinoblastoma" *Proc. Natl. Acad. Sci. U.S.A* 60: 820- (1971)
161. Knudson, A. G. Jr. "Genetic influences in human tumors" in: 'Cancer: A Comprehensive Treatise' 2nd Edition F. F. Becker (ed.) Plenum Press, N.Y. 1982 pp 73-88
162. Knudson, A. G., Jr. "Hereditary cancer, oncogenes, and antioncogenes" *Cancer Res.* 45: 1437-1443 (1985)
163. Knudson, A. G. Jr. and Meadows, A. T. "Developmental genetics of neuroblastoma" *J. Natl. Cancer Inst.* 57: 675-682 (1976)
164. Kosswig, C. "Melanotische Geschwulstbildungen bei Fischbastarden" *Verhandl. d. Deutsche Zool. Gesellsch.* 1929: 90-98 (1929)
165. Kosswig, C. "Über die veränderte Wirkung von Farbgenen in fremden Genotypen" *Biologica Generalis* 13: 276-293 (1937)
166. Koufos, A., Hansen, M. F., Copeland, N. G., Jenkins, N. A., Lampkin, B. C. and Cavanee, W. K. "Loss of heterozygosity in three embryonal tumours suggests a common pathogenetic mechanism" *Nature* 316: 330-334 (1985)
167. Koufos, A., Hansen, M. F., Lampkin, B. C., Workman, M. L., Copeland, N. G., Jenkins, N. A. and Cavanee, W. K. "Loss of alleles at loci on human chromosome 11 during genesis of Wilm's tumour" *Nature* 309: 170-172 (1984)
168. Krueger, J. G., Garber, E. A. and Goldberg, A. R. "Subcellular localization of pp60^{src} in RSV-transformed cells" *Curr. Top. Microbiol. Immunol.* 107: 52-124 (1983)
169. Krontiris, T. G. and Cooper, G. M. "Transforming activity of human tumor DNAs" *Proc. Natl. Acad. Sci. U.S.A.* 78: 1181-1184 (1981)
- 169a. Kruh, G. D., Kraus, M.H., Popescu, N.C., Amsbaugh, S.C., McBride, W.O. and Aaronson, S.A. "A novel human gene closely related to the *abl* proto-oncogene" *Science* 234: 1545-1548 (1986)
170. Kruijer, W., Cooper, J. A., Hunter, T. and Verma, I. M. "Platelet-derived growth factor induces rapid but transient expression of the *c-fos* gene and protein" *Nature* 312: 711-716 (1984)

171. Kunkel, L. M., Monaco, A. P., Middlesworth, W., Ochs, H. D. and Latt, S. A. "Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion" *Proc. Natl. Acad. Sci. U.S.A* 82: 4778-4782 (1985)
172. Land, H., Parada, L. F. and Weinberg, R. A. "Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes" *Nature* 304: 596-602 (1983)
173. Lapeyre, J. and Becker, F. F. "5-methylcytosine content of nuclear DNA during chemical hepatocarcinogenesis and in carcinomas which result" *Biochem. Biophys. Res. Commun.* 87: 698-705 (1979)
174. Lawn, I. D. "Porifera" in: 'Electrical Conduction and Behavior in Simple Invertebrates' G. A. B. Shelton (ed.) Clarendon Press, Oxford. 1982. pp 49-72.
175. Lawn, I. D., Mackie, G. O. and Silver, G. "Conduction system in a sponge" *Science* 211: 1169-1171 (1981)
176. Lentz, T. L. "Histochemical localization of neurohumors in a sponge" *J. Exp. Zool.* 162: 171-180 (1966)
177. Levinson, A. D. "Normal and activated *ras* oncogenes and their encoded products" *Trends in Genetics*, March 1986. pp 81-85
178. Li, F. P. and Fraumeni, J. F. Jr. "Prospective study of a family cancer syndrome" *J. Amer. Med. Assoc.* 247: 2692-2694 (1982)
179. Little, C. C. "The genetics of cancer in mice" *Biol. Rev.* 22: 315-343 (1947)
180. Livneh, E., Glazer, L., Segal, D., Schlessinger, J. and Shilo, B-Z. "The *Drosophila* EGF receptor homolog: conservation of both hormone binding and kinase domains" *Cell* 40: 599-607 (1985)
181. Loewenstein, W. R. "On the genesis of cellular communication" *Devel. Biol.* 15: 503-520 (1967)
182. Lowe, D. G., Capon, D. J., Delivart, E., Sakaguchi, A. Y., Naylor, S. L. and Goeddel, D. V. "Structure of the human and murine *R-ras* genes, novel genes closely related to *ras* proto-oncogenes" *Cell* 48: 137-146 (1987)
183. Lowy, D. R. and Willumsen, B. M. "The *ras* gene family" *Cancer Surveys* 5: 275-289 (1986)
184. Lynch, H. T. "Genetics, etiology and human cancer" *Prev. Med.* 9: 231-243 (1980)
185. Lynch, H. T., Schuelke, G.S. and Lynch, J.F. "Genetics of rectal cancer" *Bull. Cancer (Paris)* 71: 1-15 (1984)
186. Mackie, G. O. "Is there a conduction system in sponges?" *Coll. intern. C. N. R. S.* 291: 145-151 (1979)
187. Maeda, N. and Smithies, O. "The evolution of multigene families: human haptoglobin genes" *Ann. Rev. Genet.* 20: 81-108 (1986)
188. Mahon, K. A., Chepilinsky, A. B., Khillan, J. S., Overbeek, P. A., Piatgorsky, J. and Westphal, H. "Oncogenesis of the lens in transgenic mice" *Science* 235: 1622-1628 (1987)
- 188a. Maniatis, T., Fritsch, E. F. and Sambrook, J. "Molecular Cloning: a laboratory manual" CSH, 1982

189. March, C. J., Mosley, B., Larsen, A., Cerretti, D. P., Braedt, G., Price, V., Gillis, S., Henney, C. S., Kronheim, S. R., Grabstein, K., Conlon, P. J., Hopp, T. P. and Cosman, D. "Cloning, sequence and expression of two distinct human interleukin-1 complementary DNAs" *Nature* 315: 641-647 (1985)
190. Marchionni, M. and Gilbert, W. "The triosephosphate isomerase gene from maize: Introns antedate the plant-animal divergence" *Cell* 46: 133 (1986)
- 190a. Marks, M. A., Chipperfield, R. and Birnboim, H. C. "A procedure for the large-scale isolation of highly purified plasmid DNA using alkaline extraction and binding to glass powder" *Anal. Biochem.* 121: 382-387 (1982)
191. Marth, J. D., Peet, R., Krebs, E. G. and Perlmutter, R. M. "A lymphocyte-specific protein-tyrosine kinase gene is rearranged and overexpressed in the murine T cell lymphoma LSTRA" *Cell* 43: 393-404 (1985)
192. Martin, G. S. "Rous sarcoma virus: a function required for the maintenance of the transformed state" *Nature* 227: 1021-1023 (1970)
193. Martinez, R., Mathey-Prevot, B., Bernards, A. and Baltimore, D. "Neuronal pp60^{c-src} contains a six-amino acid insertion relative to its non-neuronal counterpart" *Science* 237: 411-415 (1987)
194. Martin-Zanca, D., Hughes, S. H. and Barbacid, M. "A human oncogene formed by the fusion of truncated tropomyosin and protein tyrosine kinase sequences" *Nature* 319: 743-748 (1986)
195. Marx, J. L. "Cell growth control takes balance" *Science* 239: 975-976 (1988)
- 195a. Matsushime, H. et al. "A possible new member of tyrosine kinase family, human FRT sequence, is highly conserved in vertebrates and located on human chromosome 13" *Jpn. J. Cancer. Res. (Gann)* 78: 655-671 (1987)
1. Mäkelä, T. P. and Allitalo, K. "Proto-oncogene amplification: Role in tumour progression" *Ann. Clin. Res.* 18: 290-296 (1986)
197. Mäueler, W., Raulf, F. and Scharf, M. "Expression of protooncogenes in embryonic, adult, and transformed tissue of *Xiphophorus* (Teleostei: Poeciliidae)" *Oncogene* 2: 421-430 (1988)
198. McKnight, G. L., O'Hara, P. J. and Parker, M. L. "Nucleotide sequence of the triosephosphate isomerase gene from *Aspergillus nidulans* : Implication for a differential loss of introns" *Cell* 46: 143-147 (1986)
199. McMenamin, M. A. S. "The emergence of animals" *Sci. Am.* 256:84-92 (1987)
200. Mehla, P. P., Bertram, J. S. and Loewenstein, W. R. "Growth inhibition of transformed cells correlates with their functional communication with normal cells" *Cell* 44: 187-196 (1986)
- 200a. Mierendorf, R. C. and Pfeiffer, D. "Direct sequencing of denatured plasmid DNA" *Meth Enzymol.* 152: 556-562 (1987)
201. Mintz, B. and Illmensee, K. "Normal genetically mosaic mice produced from malignant teratocarcinoma cells" *Proc. Natl. Acad. Sci. U.S.A* 72: 3585-3589 (1975)
202. Moloney, J. B. In: 'Conf. on Murine Leukemia' (Natl. Cancer Inst. Monograph #22)

203. Mondal, S., Brankow, D. W. and Heidelberger, C. "Two-stage chemical carcinogenesis in cultures of C3H/10T1/2 cells" *Cancer Res.* 36: 2254-2260 (1976)
204. Morizot, D. C. and Siciliano, M. J. "Linkage group V of platyfishes and swordtails of the genus *Xiphophorus* (Poeciliidae): Linkage of loci for malate dehydrogenase-2 and esterase-1 and esterase-4 with a gene controlling the severity of hybrid melanomas" *J. Natl. Cancer Inst.* 71: 809-813 (1983)
205. Moscona, A. A. "Studies of cell aggregation: demonstration of materials with selective cell binding activity" *Proc. Natl. Acad. Sci. U.S.A* 49: 742-747 (1963)
206. Mouchiroud, D., Fichant, G. and Bernardi, G. "Compositional compartmentalization and gene composition in the genome of vertebrates" *J. Mol. Evol.* 26: 198-204 (1987)
207. Mount, S. M. "A catalogue of splice junction sequences" *Nucl. Acids Res.* 10: 459-472 (1982)
208. Mulvihill, J. J. "Clinical ecogenetics of human cancer" in: 'Genes and Cancer' Bishop, J. M., Rowley, J. D. and Greaves, M. (eds.) Alan R Liss, N.Y. 1984 pp 19-36
209. Murray, M. J., Shilo, B.-Z., Shih, C., Cowing, D., Hsu, H. W. and Weinberg, R. A. "Three different human tumor cell lines contain different oncogenes" *Cell* 25: 355-361 (1981)
210. Müller, R. "Cellular and viral *fos* genes: structure, regulation of expression and biological properties of their encoded products" *Biochim. biophys. Acta* 823: 207-225 (1986)
211. Naharro, G., Robbins, K. C. and Reddy, E. P. "Gene product of *v-fgr* onc: hybrid protein containing a portion of actin and a tyrosine-specific protein kinase" *Science* 223: 63-66 (1984)
212. Neckameyer, W. S. and Wang, L.-H. "Nucleotide sequence of avian sarcoma virus UR2 and comparison of its transforming gene with other members of the tyrosine protein kinase oncogene family" *J. Virol.* 53: 879-884 (1985)
- 212a Needleman, S. B. and Wunsch, C. D. "A general method applicable to the search for similarities in the amino acid sequence of two proteins" *J. Mol. Biol.* 48: 443-453 (1970)
213. Nelson, M. and McClelland, M. "The effect of site-specific methylation on restriction-modification enzymes" *Nucl. Acids Res.* 15 (Suppl.) 1987. r219
214. 'Neoplasms and Related Disorders of Invertebrate and Lower Vertebrate Animals' *Natl. Cancer Inst. Monograph* 31, 1969.
215. Neuman-Silberberg, F. S., Schejter, E., Hoffmann, F. M. and Shilo, B. "The *Drosophila* *ras* oncogenes: structure and nucleotide sequence" *Cell* 37: 1027-1033 (1984)
216. Nishimura, S. and Sekiya, T. "Human cancer and cellular oncogenes" *Biochem. J.* 243: 313-327 (1987)
217. Nishizawa, M., Semba, K., Yoshida, M. C., Yamamoto, T., Sasaki, M and Toyoshima, K. "Structure, expression, and chromosomal location of the human *c-fgr* gene" *Molec. Cell. Biol.* 6: 511-517 (1986)
218. Noda, M., Selinger, Z., Scolnick, E. M. and Bassin, R. H. "Flat revertants isolated from Kirsten sarcoma virus-transformed cells are resistant to the action of specific oncogenes" *Proc. Natl. Acad. Sci. U.S.A* 80: 5602-5606 (1983)

219. Norris, W. *Edinburgh Med. Surg. J.* **16**: 562-565 (1820) ; referenced in Bishop, J. M. "Trends in oncogenes" *Trends In Genetics*, Sept. 1985. pp 245-249
220. Ochman, H. and Wilson, A. C. "Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes" *J. Mol. Evol.* **26**: 74-86 (1987)
221. Ohno, S. and Atkin, N. B. "Comparative values and chromosome complements of eight species of fish" *Chromosoma* **18**: 455-466 (1966)
222. Ohno, S. 'Evolution by Gene Duplication' Springer-Verlag, Heidelberg. 1970.
223. Ohno, S. "Genetic implication of karyological instability of malignant somatic cells" *Physiol. Rev.* **51**: 496-526 (1971)
224. Ohno, S. 'Protochordata, Cyclostomata and Pisces' *Animal Cytogenetics Vol. 4: Chordata 1.* Gebrüder Borntraeger, Berlin. 1974.
225. Orkin, S. H. "Reverse genetics and human disease" *Cell* **47**: 845-850 (1986)
226. Orkin, S. H., Goldman, D. S. and Sallan, S. E. "Development of homozygosity for chromosome 11p markers in Wilms' tumour" *Nature* **309**: 172-174 (1984)
- 226a Ornstein, D. L. and Kashdan, M. A. "Sequencing DNA using ³⁵S-labelling: a trouble shooting guide" *BioTechniques* **3**: 476-483 (1985)
227. Osgood, E. E. "A unifying concept of the etiology of the leukemias, lymphomas, and cancers" *J. Natl. Cancer Inst.* **18**: 155-166 (1957)
228. Oskarrson, M. K., McClements, W. L., Blair, D. S., Maizel, J. V. and VandeWoude, G. *Science* **207**: 1222-1224 (1980)
229. Oskarrson, M. K., McClements, W. L., Blair, D. S., Maizel, J. V. and VandeWoude, G. *Science* **207**: 1222-1224 (1980)
- 229a Owen, S. A. *Royal London Ophthalmol. Hosp. Rep.* **16**: 323-369 (1905)
230. Ozato, K. and Wakamatsu, Y. "Multi-step genetic regulation of oncogene expression in fish hereditary melanoma" *Differentiation* **24**: 181-190 (1983)
231. Pang, D. T., Wang, J. K. T., Valtorta, F., Benfenati, F. and Greengard, P. "Protein tyrosine phosphorylation in synaptic vesicles" *Proc. Natl. Acad. Sci. U.S.A* **85**: 762-766 (1988)
232. Parada, L. F., Tabin, C. J., Shih, C. and Weinberg, R. A. "Human EJ bladder carcinoma is a homologue of Harvey sarcoma virus ras gene" *Nature* **297**: 474-478 (1982)
233. Parker, G. H. "The reactions of sponges, with a consideration of the origin of the nervous system" *J. Exp. Zool.* **8**: 1-41 (1910)
234. Parker, R. C., Varmus, H. E. and Bishop, J. M. "Cellular homologue (c-src) of the transforming gene of Rous sarcoma virus: Isolation, mapping, and transcriptional analysis of c-src and flanking regions" *Proc. Natl. Acad. Sci. U.S.A* **78**: 5842-5846 (1981)
235. Parker, R. C., Varmus, H. E. and Bishop, J. M. "Expression of v-src and chicken c-src in rat cells demonstrates qualitative differences between pp60^{v-src} and pp60^{c-src}" *Cell* **37**: 131-139 (1984)
236. Parnaik, V. K., Patil, D. N. and Das, M. R. "Undermethylation of the c-myc gene in a rat

- hepatoma correlates with increased expression of the gene" *Curr. Sci.* 56: 989-994 (1987)
237. Patel, R. S., Odermatt, E., Schwarzbauer, J. E. and Hynes, R. O. "Organization of the fibronectin gene provides evidence for exon shuffling during evolution" *EMBO J.* 6: 2565-2572 (1987)
 238. Pavans de Ceccatty, M. "Le systeme nerveux des eponges calcaires et siliceuses" *Ann. Sci. Nat. (Zool.)* 17: 203-290 (1955)
 239. Pavans de Ceccatty, M. "Coordination in sponges. The foundations of integration" *Am. Zool.* 14: 895-903 (1974)
 240. Pavans de Ceccatty, M. "The origin of the integrative systems: a change in view derived from research on coelenterates and sponges" *Perspec. Biol. Med.* 17: 379-390 (1974)
 - 240a Pfeiffer, D. and Mierendorf, R. "Direct sequencing of pGEM plasmids: Klenow vs. reverse transcriptase" *Promega Notes* 5: 1-4 (1986)
 241. Piwnicka-Worms, H., Saunders, K. B., Roberts, T. M., Smith, A. E. and Cheng, S. H. "Tyrosine phosphorylation regulates the biochemical and biological properties of pp60^{c-src}" *Cell* 49: 75-82 (1987)
 242. Podesta, A. H., Mullins, J., Pierce, G. B. and Wells, R. S. "The neurula stage mouse embryo in control of neuroblastoma" *Proc. Natl. Acad. Sci. U.S.A* 81: 7608-7611 (1984)
 243. Quintrell, N., Lebo, R., Varmus, H. E., Bishop, J. M., Pettenati, M. J., Le Beau, M. M., Diaz, M. O. and Rowley, J. D. "Identification of a human gene (HCK) that encodes a protein-tyrosine kinase and is expressed in hemopoietic cells" *Molec. Cell. Biol.* 7: 2267-2275 (1987)
 244. Radany, E. W. "Isolation and expression of genes with homology to the tyrosine kinase domain of viral oncogenes during sea urchin embryogenesis" *ISCU Short Reports* 7: 94 (1987)
 - 244a Ralston, R. and Bishop, J.M. "The product of the protooncogene *c-src* is modified during the cellular response to platelet-derived growth factor" *Proc. Natl. Acad. Sci. U.S.A.* 82: 7845-7849 (1985)
 245. Rasmont, R. "Freshwater sponges as a material for the study of cell differentiation" *Curr. Top. Devel. Biol.* 10: 141-159 (1975)
 246. Raulf, F., Mäueler, W., Robertson, S. M. and Scharf, M. "Cellular localization of proto-oncogene mRNA in *Xiphophorus* : Expression of *c-src* in differentiated bipolar neurons of the adult neural retina" (Submitted for publication)
 247. Reddy, E. P., Smith, M. J., Canaan, E., Robbins, K. C., Tronick, S. R., Zain, S. and Aaronson, S. A. "Nucleotide sequence analysis of the transforming region and large terminal redundancies of Moloney murine sarcoma virus" *Proc. Natl. Acad. Sci. U.S.A* 77: 5234-5238 (1980)
 - 247a Reed, K. C. and Mann, D. A. "Rapid transfer of DNA from agarose gels to nylon membranes" *Nucl. Acids Res.* 13: 7207-7221 (1985)
 248. Reymond, C. D., Gomer, R. H., Mehdy, M. C. and Firtel, R. A. "Developmental regulation of a *Dictyostelium* gene encoding a protein homologous to mammalian *ras* protein" *Cell* 39: 141-148 (1984)
 249. Ribbert, H. "Über die Histogenese und das Wachstum des Carcinoms" *Arch. path. Anat.* 141: 153-177 (1895) ; referenced in: Furth, J. "Conditioned and autonomous neoplasms: a review" *Cancer Res.* 13: 477-492 (1953)

250. Riehl, R., Scharff, M. and Kollinger, G. "Comparative studies on the ultrastructure of malignant melanoma in fish and human by freeze-etching and transmission electron microscopy" *J. Cancer Res. Clin. Oncol.* 107: 21-31 (1984)
- 250a. Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, P. "Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I" *J. Mol. Biol.* 113: 237-251 (1977)
251. Riggs, A. D. and Jones, P. A. "5-methylcytosine, gene regulation, and cancer" *Adv. Cancer Res.* 40: 1-30 (1983)
252. Roebroek, A. J. M., Schalken, J. A., Onnekink, C., Bloemers, H. P. J. and Van de Ven, W. J. M. "Structure of the feline *c-fes/lfs* proto-oncogene: genesis of a retroviral oncogene" *J. Virol.* 61: 2009-2016 (1987)
253. Rohrschneider, L. R. and Gentry, L. E. "Subcellular locations of retroviral transforming proteins define multiple mechanisms of transformation" *Adv. Viral Oncol.* 4: 269-306 (1984)
254. Roitt, I. M., Brostoff, J. and Male, D. K. "Evolution of immunity" in: 'Immunology' Gower Medical Publishing, London, 1985. pp 15.1-15.12
255. Rosen, D. E. "Fishes from uplands and inter-mountain basins of Guatemala: Revisionary studies and comparative geography" *Bull. Am. Mus. Natl. Hist.* 162: 267-376 (1979)
256. Rowley, J. D. "Biological implications of consistent chromosome rearrangements in leukemia and lymphoma" *Cancer Res.* 44: 3159-3168 (1984)
257. Rubin, C. S. and Rosen, O. M. "Protein phosphorylation" *Ann. Rev. Biochem.* 44: 831-887 (1975)
258. Rülter, U., Garber, C., Komitowski, D., Müller, R. and Wagner, E. F. "Deregulated *c-fos* expression interferes with normal bone development in transgenic mice" *Nature* 325: 412-416 (1987)
259. Sager, R. "Genetic suppression of tumour formation: a new frontier in cancer research" *Cancer Res.* 46: 1573-1580 (1986)
260. Sandberg, M., Vuorio, T., Hirvonen, H., Alitalo, K. and Vuorio, E. "Enhanced expression of TGF- β and *c-fos* mRNAs in the growth plates of developing human long bones" *Development* 102: 461-470 (1988)
- 260a. Sanger, F., Nicklen, S. and Coulson, A. P. "DNA sequencing with chain-terminating inhibitors" *Proc. Natl. Acad. Sci. U.S.A.* 74: 5463-5467 (1977)
261. Scharff, M., Scharff, A. and Anders, F. "Phenotypic conversion of malignant melanomas to benign melanoma and vice versa in *Xiphophorus*" *Pigment Cell* 1981; Proceedings of the XIth International Pigment Cell Conference, Sendai, Japan. M. Seiji (ed.) University of Tokyo Press, 1981. pp 507-514
262. Scharff, M. and Barnekow, A. "The expression in eukaryotes of a tyrosine kinase which is reactive with pp60 *c-src* antibodies" *Differentiation* 23: 109-114 (1982)
263. Scharff, M., Barnekow, A., Bauer, H. and Anders, F. "Correlations of inheritance and expression between a tumor gene and the cellular homolog of the Rous sarcoma virus-transforming gene in *Xiphophorus*" *Cancer Res.* 42: 4222-4227 (1982)

264. Scharf, M. and Barnekow, A. "Differential expression of the cellular *src* gene during vertebrate development" *Dev. Biol.* 105: 415-422 (1984)
- 264a. Scharf, M., Schmidt, C.-R., Anders, A. and Barnekow, A. "Elevated expression of the cellular *src* gene in tumors of differing etiologies in *Xiphophorus* " *Int. J. Cancer* 36: 199-207 (1985)
265. Scharf, M. "A sex-chromosomal restriction fragment length marker linked to melanoma determining loci (*Tu*) in *Xiphophorus* " *Genetics* 119: (in press) 1988
266. Scharf, M., Holstein, T., Robertson, S. M., David, C. N. and Barnekow, A. (in preparation)
267. Schwab, M. "Genome organization in *Xiphophorus* (Poeciliidae; Teleostei)" *Mol. Gen. Genet.* 188: 410-417 (1982)
268. Schwab, M. "Malignant metamorphosis: Developmental genes as culprits for oncogenesis in *Xiphophorus* " *Adv. Cancer Res.* 47: 63-97 (1986)
269. Schwab, M. "Oncogenes and tumor suppressor genes in *Xiphophorus* " *Trends Genet.* 3: 38-42 (1987)
270. Schwartz, D. E., Tizard, R. and Gilbert, W. "Nucleotide sequence of Rous sarcoma virus" *Cell* 32: 853-869 (1983)
271. Sefton, B. M. "Oncogenes encoding protein kinases" *Trends in Genetics*, Nov. 1985. pp 306-308
272. Sefton, B. M. "The viral tyrosine protein kinases" *Curr. Topics Micro. Immunol.* 123: 39-72 (1986)
273. Seizinger, B. R., Rouleau, G., Ozelius, L. J., Lane, A. H., St George-Hyslop, P., Huson, S., Gusella, J. F. and Martuza, R. L. "Common pathogenetic mechanism for three tumor types in bilateral acoustic neurofibromatosis" *Science* 236: 317-319 (1987)
274. Seki, T., Fujii, G., Mori, S., Tamaoki, N. and Shibuya, M. *Jpn. J. Cancer Res. (Gann)* 76: 907-910 (1985)
275. Semba, K., Nishizawa, M., Miyajima, N., Yoshida, M. C., Sukegawa, J., Yamanashi, Y., Sasaki, M., Yamamoto, T. and Toyoshima, K. "*yes*-related protooncogene, *syn*, belongs to the protein-tyrosine kinase family" *Proc. Natl. Acad. Sci. U.S.A.* 83: 5459-5463 (1986)
276. Shalloway, D., Coussens, P. M. and Yaciuk, P. "Overexpression of the *c-src* protein does not induce transformation of NIH 3T3 cells" *Proc. Natl. Acad. Sci. U.S.A.* 81: 7071-7075 (1984)
277. Sherr, C. J., Fedele, L. A., Oskarsson, M., Maizel, J. and Vande Woude, G. "Molecular cloning of Snyder-Theilen feline leukemia and sarcoma viruses: Comparative studies of feline sarcoma virus with its natural helper virus and with Moloney murine sarcoma virus" *J. Virol.* 34: 200-212 (1980)
278. Shih, C., Shilo, B. Z., Goldfarb, M. P., Dannenberg, A. and Weinberg, R. A. "Passage of phenotypes of chemically transformed cells via transfection of DNA and chromatin" *Proc. Natl. Acad. Sci. U.S.A.* 76: 5714-5718 (1979)
279. Shilo, B.-Z. "Protooncogenes in *Drosophila melanogaster*" *Trends in Genetics* 3: 69-72 (1987)

280. Shimizu, K., Goldfarb, M., Suard, Y., Perucho, M., Li, Y., Kamata, T., Feramisco, J., Stayneser, E., Fogh, J. and Wigler, M. "Three human transforming genes are related to the viral *ras* oncogene" *Proc. Natl. Acad. Sci. U.S.A* 80: 2112-2116 (1983)
281. Siciliano, M. J., Morizot, D. C. and Wright, D. A. "Factors responsible for platyfish-swordtail hybrid melanoma - many or few?" *Pigment Cell Vol. 2*. Karger, Basel. 1976. pp 47-58
282. Siciliano, M. J. and Wright, D. A. "Biochemical genetics of the platyfish-swordtail hybrid melanoma system" *Prog. Exp. Tumor Res.* 20: 398-411 (1976)
283. Simon, M. A., Drees, B., Kornberg, T. and Bishop, J. M. "The nucleotide sequence and the tissue-specific expression of *Drosophila* *c-src*" *Cell* 42: 831-840 (1985)
- 283a. Smith, T. F. and Waterman, M. S. "Comparison of Bio-sequences" *Adv. Appl. Math.* 2: 482-489 (1981)
284. Sobel, H. J., Marquet, E., Kallman, K. D. and Corley, G. J. "Melanomas in platy/swordtail hybrids" In: 'The Pathology of Fishes' W. E. Ribelin and G. Migaki (eds.) The University of Wisconsin Press, Madison, 1975. pp 945-981
285. Sorge, J. P., Levy, B. T. and Maness, P. F. "pp60^{c-src} is developmentally regulated in the neural retina" *Cell* 36: 249-257 (1984)
286. Sparkes, R. S., Murphree, A. L., Lingua, R. W., Sparkes, M. C., Field, L. L., Funderburk, S. J. and Benedict, W. F. "Gene for hereditary retinoblastoma assigned to human chromosome 13 by linkage to esterase D" *Science* 219: 971-973 (1983)
287. Spencer, A. N. and Schwab, W. E. "Hydrozoa" In: 'Electrical Conduction and behavior in simple invertebrates' G. A. B. Shelton (ed.) Clarendon Press, Oxford. 1982. pp 73-148
288. Sporn, M. B. and Roberts, A. B. "Autocrine growth factors and cancer" *Nature* 313: 745-747 (1985)
289. Sporn, M. B. and Roberts, A. B. "Peptide growth factors are multifunctional" *Nature* 332: 217-219 (1988)
290. Srivatsan, E. S., Benedict, W. F. and Stanbridge, E. J. "Chromosome 11 is implicated in the suppression of neoplastic expression in human cell hybrids" *Cancer Res.* 46: 6174-6179 (1986)
291. Stanbridge, E. J. "Suppression of malignancy in human cells" *Nature* 260: 17-20 (1976)
292. Stanbridge, E. J. "A case for human tumor-suppressor genes" *Bioessays* 3: 252-256 (1985)
293. Stanbridge, E. J., Flandermeyer, R. R., Daniels, D. W. and Nelson-Rees, W. A. "Specific chromosome loss associated with the expression of tumorigenicity in human cell hybrids" *Somatic. Cell Genet.* 7: 699-712 (1981)
294. Stanbridge, E. J., Der, C. J., Doersen, C.-J., Nishimi, R. Y., Peehl, D. M., Weissman, B. E. and Wilkinson, J. E. "Human cell hybrids: analysis of transformation and tumorigenicity" *Science* 215: 252-259 (1982)
295. Steele, R. E. "Two divergent cellular *src* genes are expressed in *Xenopus laevis*" *Nucl. Acids Res.* 13: 1747-1761 (1985)
296. Stehelin, D., Guntaka, R. V., Varmus, H. E. and Bishop, J. M. "Purification of DNA

- complementary to nucleotide sequences required for neoplastic transformation of fibroblasts by avian sarcoma viruses" *J. Mol. Biol.* 101: 349-365 (1976)
297. Stehelin, D., Varmus, H. E., Bishop, J. M. and Vogt, P. K. "DNA related to the transforming gene(s) of avian sarcoma virus is present in normal avian DNA" *Nature* 260: 170-173 (1976)
 298. Stevens, L. "The development of transplantable teratocarcinomas from intra-testicular grafts of pre- and post implantation mouse embryos" *Dev. Biol.* 21: 364-382 (1970)
 299. Stewart, T. A., Pattengale, P. K. and Leder, P. "Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/*myc* fusion genes" *Cell* 38: 627-637 (1984)
 300. Stoker, M. G. P., Shearer, M. and O'Neill, C. "Growth inhibition of polyoma-transformed cells by contact with static normal fibroblasts" *J. Cell Sci.* 1: 297-310 (1966)
 301. Strebhardt, K., Mullins, J. I., Bruck, C. and Rübsem-Waigmann, H. "Additional member of the protein-tyrosine kinase family: the *src* and *lck* -related protooncogene *c-tkl*" *Proc. Natl. Acad. Sci. U.S.A.* 84: 8778-8782 (1987)
 302. Strong, L. C. "Genetic environmental interactions" in: 'Cancer Epidemiology and Prevention' D. Schottenfeld and J. F. Fraumeni (eds) W. B. Saunders, N.Y. 1982 pp 506-516
 303. Sukegawa, J., Semba, K., Yamanashi, Y., Nishizawa, M., Miyajima, N., Yamamoto, T. and Toyoshima, K. "Characterization of cDNA clones for the human *c-yes* gene" *Molec. Cell. Biol.* 7: 41-47 (1987)
 304. Südhof, T. C., Goldstein, J. L., Brown, M. S. and Russell, D. W. "The LDL receptor gene: A mosaic of exons shared with different proteins" *Science* 228: 815-822 (1985)
 305. Takeya, T. and Hanafusa, H. "Structure and sequence of the cellular gene homologous to the RSV *src* gene and the mechanism for generating the transforming virus" *Cell* 32: 881-890 (1983)
 306. Takeya, T., Feldman, R. A. and Hanafusa, H. "DNA sequence of the viral and cellular *src* gene of chickens. I. Complete nucleotide sequence of an Eco RI fragment of recovered avian sarcoma virus which codes for gp37 and pp60^{src}" *J. Virol.* 44: 1-11 (1983)
 307. Tanaka, A., Gibbs, C. P., Arthur, R. R., Anderson, S. K. Kung, H.-J. and Fujita, D. J. "DNA sequence encoding the amino-terminal region of the human *c-src* protein: Implications of sequence divergence among *src* -type kinase oncogenes" *Molec. Cell. Biol.* 7: 1978-1983 (1987)
 308. Tavalga, W. N. "Embryonic development of the platyfish (*Platypoecilus*), the swordtail (*Xiphophorus*) and their hybrids" *Bull. Am. Mus. Natl. Hist.* 94: 163-229 (1949)
 309. Todaro, G. J. and Huebner, R. J. "The viral oncogene hypothesis: new evidence" *Proc. Natl. Acad. Sci. U.S.A.* 69: 1009-1015 (1972)
 310. Tuzet, O. "The phylogeny of sponges according to embryological, histological, and serological data, and their affinities with the protozoa and the Cnidaria" in: 'The Lower Metazoa. Comparative Biology and Phylogeny' E. C. Dougherty (ed.) Univ. Calif. Press, Berkeley. 1963. pp 129-148
 311. Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Libermann, T. A., Schlessinger, J., Downward, J., Mayes, E. L. V., Whittle, N., Waterfield, M. D. and Seeburg, P. H. "Human epidermal growth factor receptor cDNA sequence and aberrant

- expression of the amplified gene in A431 epidermoid carcinoma cells" *Nature* 309: 418-425 (1984)
312. Van Beveren, C. and Verma, I. M. "Homology among oncogenes" *Curr. Topics Micro.Immunol.* 123: 73-98 (1986)
 313. Van de Vyver, G. "Phenomena of cellular recognition in sponges" *Curr. Top. Devel. Biol.* 10: 123-140 (1975)
 314. Varmus, H. E. "The molecular genetics of cellular oncogenes" *Ann. Rev. Genetics* 18: 553- 612 (1984)
 315. Vielkind, J. and Vielkind, U. "Melanoma formation in fish of the genus *Xiphophorus* : a genetically-based disorder in the determination and differentiation of a specific pigment cell" *Can. J. Genet. Cytol.* 24: 133-149 (1982)
 316. Vielkind, J., Haas-Andela, H., Vielkind, U. and Anders, F. "The induction of a specific pigment cell type by total genomic DNA injected into the neural crest region of fish embryos of the genus *Xiphophorus* " *Mol. Gen. Genet.* 185: 379-389 (1982)
 317. Vielkind, J. R. and Dippel, E. "Oncogene-related sequences in xiphophorine fish prone to hereditary melanoma formation" *Can J. Genet. Cytol.* 26: 607-614 (1984)
 318. Vielkind, U. "Genetic control of cell differentiation in platyfish-swordtail melanomas" *J. Exp. Zool.* 196: 197-204 (1976)
 319. Voronova, A. F. and Sefton, B. M. "Expression of a new tyrosine protein kinase is stimulated by retroviral promoter insertion" *Nature* 319: 682-685 (1986)
 320. Waalwijk, C. and Flavell, R. A. "DNA methylation at a CCGG sequence in the large intron of the rabbit β -globin gene: tissue specific variations" *Nucl. Acids Res.* 5:4631-4641(1978)
 321. Wang, J. L. and Hsu, Y-M. "Negative regulators of cell growth" *TIBS* 11: 24-26 (1986)
 322. Wang, J. Y. J., Ledley, F., Goff, S., Lee, R., Groner, Y. and Baltimore, D. "The mouse *c-abl* locus: molecular cloning and characterization" *Cell* 36: 349-356 (1984)
 323. Weber, M. J. "Malignant transformation by Rous sarcoma virus: from phosphorylation to phenotype" In: *Advances in Viral Oncology*, Vol. 4 G. Klein (ed.) Raven Press, N.Y. 1984 pp 249-268
 324. Wegener, H. "Klonierung speziesspezifischer repetitiver Sequenzen von *Xiphophorus* " Diplomarbeit, Ludwig-Maximilians-Universität, Munich. 1987.
 325. Weissman, B. E., Saxon, P. J., Pasquale, S. R., Jones, G. R., Geiser, A. G. and Stanbridge, E. J. "Introduction of a normal human chromosome 11 into a Wilm's tumor cell line controls its tumorigenic expression" *Science* 236: 175-180 (1987)
 326. Wells, R. S. and Miotto, K. A. "Widespread inhibition of neuroblastoma cells in the 13- to 17-day-old mouse embryo" *Cancer Res.* 46: 1659-1662 (1986)
 327. Wigler, M., Levy, D. and Perucho, M. "The somatic replication of DNA methylation" *Cell* 24: 33-40 (1981)
 - 327a. Williams, S. A. et al *BioTechniques* 4: 138-147 (1986)
 328. Wilson, A. C., Carlson, S. S. and White, T. J "Biochemical evolution" *Ann. Rev. Biochem*

46: 573-639 (1977)

329. Wilson, A. C., Ochman, H. and Prager, E. M. "Molecular time scale for evolution" *Trends Genet.* 3: 241-247 (1987)
330. Wittbrodt, J. "Versuche zur Isolierung und Klonierung von molekularen Markersequenzen für einen Melanom induzierenden Locus (Tu) bei *Xiphophorus*" Diplomarbeit, Ludwig-Maximilians-Universität, Munich. 1987.
331. Woolford, J. and Beemon, K. "Transforming proteins of Fujinami and PRCII avian sarcoma viruses have different subcellular locations" *Virology* 135: 168-180 (1984)
332. Wu, C-I. and Li, W-H. "Evidence for higher rates of nucleotide substitution in rodents than in man" *Proc. Natl. Acad. Sci. U.S.A* 82: 1741-1745 (1985)
333. Yamamoto, T., Nishida, T., Miyajima, N., Kawai, S., Ooi, T. and Toyoshima, K. "The *erb B* gene of avian erythroblastosis virus is a member of the *src* gene family" *Cell* 35: 71-78 (1983)
334. Yamanashi, Y., Fukushima, S-I., Semba, K., Sukegawa, J., Miyajima, N., Matsubara, K-I., Yamamoto, T and Toyoshima, K. "The *yes* -related cellular gene *lyn* encodes a possible tyrosine kinase similar to *p56^{lck}*" *Molec. Cell. Biol.* 7: 237-243 (1987)
- 334a Yanisch-Perron, C., Viera, J. and Messing, J. "Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors" *Gene* 33: 103-119 (1985)
335. Yisraeli, J. and Szyf, M. "Gene methylation patterns and expression" in: "DNA methylation: biochemistry and biological significance" Razin, A., Cedar, H. and Riggs, A. D. (eds.) Springer-Verlag, N. Y. 1984. pp 353-378
336. Yisraeli, J., Adelstein, R. S., Melloul, D., Nudel, U., Yaffe, D. and Cedar, H. "Muscle-specific activation of a methylated chimeric actin gene" *Cell* 46: 409-416 (1986)
337. Yunis, J. J. "The chromosomal basis of human neoplasia" *Science* 221: 227-236 (1983)
338. Ziegler, S. F., Marth, J. D., Lewis, D. B. and Perlmutter, R. M. "Novel protein-tyrosine kinase gene (*hck*) preferentially expressed in cells of hematopoietic origin" *Molec. Cell. Biol.* 7: 2276-2285 (1987)

APPENDIX 1

CHICKEN C-SRC

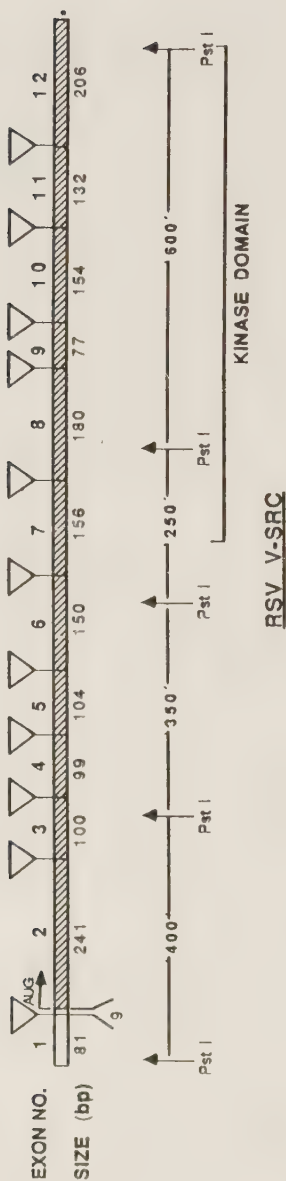


FIGURE 1 Comparison of the chicken c-src and Rous sarcoma virus v-src genes. For the chicken c-src gene the exons are numbered above and sizes of the exons (in basepairs) are given below. For the RSV v-src gene the location of the Pst I sites are indicated that give rise to the four v-src specific fragments that have been used throughout the work presented in this dissertation.

Appendix 2 CODON USAGE

		HUMCSRC		MUSCSRC		CHKCSRC		XIPHCSRC	
GLY	GGG	16.76	0.21	26.17	0.33	24.34	0.32	12.93	0.17
GLY	GGA	11.17	0.14	11.21	0.14	14.98	0.20	21.55	0.28
GLY	GGT	9.31	0.12	11.21	0.14	5.62	0.08	12.93	0.17
GLY	GGC	42.83	0.53	29.91	0.38	29.96	0.40	30.17	0.39
GLU	GAG	72.63	0.93	61.68	0.79	59.93	0.80	56.03	0.81
GLU	GGA	5.59	0.07	16.82	0.21	14.98	0.20	12.93	0.19
ASP	GGT	5.59	0.14	5.61	0.14	7.49	0.17	21.55	0.56
ASP	GGC	33.52	0.86	33.64	0.86	35.58	0.83	17.24	0.44
VAL	GTG	42.83	0.77	41.12	0.71	31.84	0.57	25.86	0.33
VAL	GTA	0.00	0.00	5.61	0.10	0.00	0.00	4.31	0.06
VAL	GTT	1.86	0.03	3.74	0.06	5.62	0.10	30.17	0.39
VAL	GTC	11.17	0.20	7.48	0.13	18.73	0.33	17.24	0.22
ALA	GCG	11.17	0.14	7.48	0.10	5.62	0.09	21.55	0.28
ALA	GCA	3.72	0.05	5.61	0.07	13.11	0.20	0.00	0.00
ALA	GCT	16.76	0.20	16.82	0.22	14.98	0.23	21.55	0.28
ALA	GCC	50.28	0.61	46.73	0.61	31.84	0.49	34.48	0.44
ARG	AGG	7.45	0.13	9.35	0.16	7.49	0.12	8.62	0.13
ARG	AGA	3.72	0.05	5.61	0.09	1.87	0.03	21.55	0.33
SER	AGT	3.72	0.05	3.74	0.05	1.87	0.03	0.00	0.00
SER	AGC	20.48	0.30	18.69	0.26	24.34	0.35	4.31	0.11
LYS	AAG	46.55	0.81	39.25	0.68	50.56	0.90	43.10	0.71
LYS	AAA	11.17	0.19	18.69	0.32	5.62	0.10	17.24	0.29
ASN	AAT	3.72	0.11	7.48	0.21	1.87	0.06	8.62	0.40
ASN	AAC	29.80	0.89	28.04	0.79	31.84	0.94	12.93	0.60
MET	ATG	20.48	1.00	20.56	1.00	22.47	1.00	25.86	1.00
ILE	ATA	0.00	0.00	1.87	0.06	1.87	0.06	4.31	0.08
ILE	ATT	5.59	0.19	7.48	0.25	3.75	0.13	12.93	0.25
ILE	ATC	24.21	0.81	20.56	0.69	24.34	0.81	34.48	0.67
THR	ACG	14.90	0.22	7.48	0.11	13.11	0.17	12.93	0.21
THR	ACA	11.17	0.17	13.08	0.19	13.11	0.17	17.24	0.29
THR	ACT	3.72	0.06	11.21	0.17	16.85	0.22	12.93	0.21
THR	ACC	37.24	0.56	35.51	0.53	33.71	0.44	17.24	0.29
TRP	TGG	16.76	1.00	16.82	1.00	16.85	1.00	21.55	1.00
END	TGA	---	---	---	---	---	---	---	---
CYS	TGT	1.86	0.11	5.61	0.33	0.00	0.00	4.31	0.25
CYS	TGC	14.90	0.89	11.21	0.67	16.85	1.00	12.93	0.75
END	TAG	---	---	---	---	---	---	---	---
END	TAA	---	---	---	---	---	---	---	---

TYR	TAT	9.31	0.22	14.95	0.35	7.49	0.17	12.93	0.38
TYR	TAC	33.52	0.78	28.04	0.65	35.58	0.83	21.55	0.63
LEU	TTG	3.72	0.04	5.61	0.06	5.62	0.06	8.62	0.07
LEU	TTA	1.86	0.02	0.00	0.00	0.00	0.00	8.62	0.07
PHE	TTT	9.31	0.24	11.21	0.29	11.24	0.29	21.55	0.63
PHE	TTC	29.80	0.76	28.04	0.71	28.09	0.71	12.93	0.38
SER	TCG	11.17	0.16	9.35	0.13	13.11	0.19	0.00	0.00
SER	TGA	7.45	0.11	7.48	0.11	3.75	0.05	12.93	0.33
SER	TGT	5.59	0.08	3.74	0.05	3.75	0.05	12.93	0.33
SER	TGC	20.46	0.30	28.04	0.39	22.47	0.32	8.62	0.22
ARG	CGG	29.80	0.50	29.91	0.50	29.96	0.47	0.00	0.00
ARG	CGA	1.86	0.03	1.87	0.03	1.87	0.03	12.93	0.20
ARG	CGT	1.86	0.03	1.87	0.03	3.75	0.06	0.00	0.00
ARG	CGC	14.90	0.25	11.21	0.19	18.73	0.29	21.55	0.33
GLN	CAG	40.97	0.96	39.25	0.91	39.33	0.91	38.79	0.82
GLN	CAA	1.86	0.04	3.74	0.09	3.75	0.09	8.62	0.18
HIS	CAT	1.86	0.11	3.74	0.22	7.49	0.40	4.31	0.25
HIS	CAC	14.90	0.89	13.08	0.78	11.24	0.60	12.93	0.75
LEU	CTG	57.73	0.63	56.07	0.61	59.93	0.65	64.66	0.56
LEU	CTA	0.00	0.00	7.48	0.08	1.87	0.02	4.31	0.04
LEU	CTT	1.86	0.02	3.74	0.04	0.00	0.00	4.31	0.04
LEU	CTC	26.07	0.29	18.69	0.20	24.34	0.27	25.86	0.22
PRO	CCG	11.17	0.19	9.35	0.16	5.62	0.09	8.62	0.18
PRO	CCA	5.59	0.09	5.61	0.09	9.36	0.15	8.62	0.18
PRO	CCT	9.31	0.18	7.48	0.13	3.75	0.06	12.93	0.27
PRO	COC	33.52	0.56	37.38	0.63	43.07	0.70	17.24	0.36

3rd BASE OF CODONS

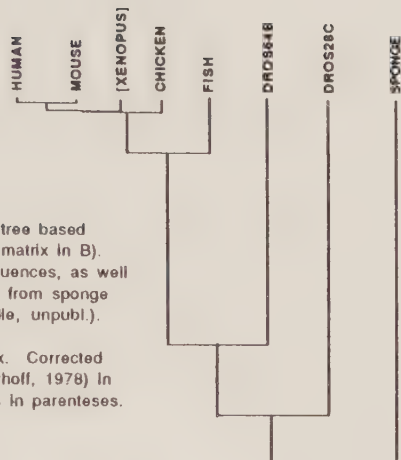
HUMAN C-SRC

G	0.40
A	0.07
T	0.09
C	0.44
G + C	0.84

XIPHOPHORUS C-SRC

G	0.35
A	0.16
T	0.19
C	0.30
G + C	0.65

A)



APPENDIX 3

A) An expanded phylogenetic tree based upon the amino acid distance matrix in B). Included are the two Dsrc sequences, as well as some preliminary sequence from sponge src-related clone 4-1 (S. Otitile, unpubl.).

B) Amino acid distance matrix. Corrected PAM values (according to Dayhoff, 1978) in bold type. Uncorrected values in parentheses.

B)

	HUMAN	MOUSE	[XENOPUS]	CHICKEN	FISH	DROS 64B	DROS 28C	SPONGE
HUMAN								
MOUSE	0.44							
[XENOPUS]	1.85	1.85						
CHICKEN	1.77	2.21	1.85					
FISH	10.0 (9.5)	11.0 (10.3)	8.5 (8.0)	11.0 (10.3)				
DROS 64B	70.5 (46.2)	70.5 (46.2)	61.5 (42.6)	70.5 (46.1)	60.0 (41.8)			
DROS 28C	106.5 (58.4)	106.5 (58.4)	108.0 (59.3)	107.0 (58.6)	88.5 (52.6)	115.0 (61.8)		
SPONGE	131.0 (64.7)	131.0 (64.7)	108.5 (58.8)	131.0 (64.7)	146.5 (67.6)	222.0 (77.9)	195.0 (75.0)	

APPENDIX 4

	SRC	YES	FYN	FGR	HCK	LYN	MOUSE LCK	CHICKEN TKL
SRC								
YES	17.2 (8.6)							
FYN	26.5 (14.2)	19.9 (12.1)						
FGR	31.0 (17.5)	29.0 (15.7)	26.4 (16.6)					
HCK	50.7 (30.0)	43.7 (27.2)	46.8 (27.2)	51.0 (31.7)				
LYN	56.6 (29.1)	52.0 (25.2)	50.3 (28.0)	54.3 (31.1)	26.6 (11.8)			
MOUSE LCK	57.2 (33.8)	47.5 (29.0)	50.8 (29.5)	58.8 (33.9)	34.0 (19.0)	36.0 (20.4)		
CHICKEN TKL	51.4 (31.0)	51.0 (29.2)	48.1 (26.9)	57.0 (32.8)	32.5 (18.2)	37.7 (19.5)	17.9 (10.0)	

FIGURE 1 Amino acid distance matrix for the known members of the src gene-family. Values are corrected according to Dayhoff, 1978 (uncorrected values in parentheses). This matrix was used in generating Figure 29.

APPENDIX 5

CLONING AND INITIAL CHARACTERIZATION OF HOMEOBOX-CONTAINING GENES FROM THE TELEOST FISH *XIPHOPHORUS MACULATUS*

Both the larval and adult bodies of *Drosophila* are composed of a series of metameric segments [1]. In *Drosophila* a number of genes have been identified which are required for the establishment of this body plan. These genes fall into several classes. One class, the 'maternal effect genes', control the gross structure and polarity of the organism [13]. These genes are active during oogenesis and their products act to specify positional values in the developing zygote. A second class, the 'segmentation' genes, are transcribed in the zygote and affect the number or polarity of the segments [14]. For example, mutants in the segmentation gene *fushi tarazu* give rise to a larva with only half the normal number of segments because each alternate segment is deleted [19]. The third class of genes are the 'homeotic' genes, also expressed in the zygote and involved in the specification of segment identity [4]. For example, certain mutations in the *Bithorax* gene complex produce a fly possessing the normal number of body segments, but with the third thoracic segment transformed into a second thoracic segment, resulting in a fly with four wings [10].

The homeobox is a highly conserved DNA segment of approximately 180 basepairs that was originally discovered as a region of cross-homology between the *Drosophila* genes *Antp*, *ftz* and *Ubx* but has since been detected in over 12 *Drosophila* genes involved in determining the spatial organization of the embryo, as well as in a number of other organisms including man [4,11,15,17]. All homeoboxes sequenced to date encode a highly conserved protein domain designated the 'homeodomain' with sequence-specific DNA-binding activity. The sequence identity between the homeobox-containing genes in *Drosophila* ranges between 48-81% and 38-88% at the DNA and amino acid levels respectively [4]. The evolutionary conservation of homeobox-containing genes is extremely high - for example, the *Drosophila Antp* and the human C1 homeoboxes share 86% and 98% identity at the DNA and amino acid levels respectively [4]. The *Drosophila* genes located within the ANT-C and BX-C gene complexes (i.e. including *Antp*, *ftz* and *Ubx*) are more closely related to one another with respect to their homeobox sequences than are the homeobox-containing genes found outside these clusters [4].

The observation of homeobox-containing genes in vertebrates, including man, raises several questions, including whether these genes are homologous to the *Drosophila* genes, and whether they are involved in any aspect of pattern formation or segmentation. The sequence relationship of the mouse *engrailed* -related gene extends beyond the confines of the homeodomain, suggesting that the mammalian genome does contain a gene homologous to the *Drosophila* gene [7,8]. There is also evidence for tissue-specific expression of homeobox-containing genes in the mouse - at least two mouse genes that contain homeoboxes are differentially expressed in the central nervous system, similar to *Antp* and *Ubx* in the adult *Drosophila* [16,18]. However, it is

not known whether these genes have an analogous function to the *Drosophila* genes in earlier embryonic development - i.e. the determination of segmental number and/or identity.

Segmentation in the vertebrate embryo is most obvious in the repeating pattern of the somites, and this is reflected in the adult by the serial arrangement of the vertebrae and their associated muscles, nerves, ribs, and blood vessels [6,9]. Analogous to the situation in *Drosophila*, these embryonic segments can be functionally sub-divided into anterior and posterior halves [9]. The fish offers several advantages over the mouse for an analysis of the role(s) of homeobox-containing genes during vertebrate embryonic development: 1) whereas in *Drosophila* the segmentation is clearly both mesodermal and ectodermal and persists throughout the life of the animal, in mammals this is primarily a feature only of the mesodermal somites in the embryo; in the fish on the other hand, the segmented pattern includes all mesoderm-derived tissues (although still not apparently the ectoderm) and is maintained for certain tissues throughout the life of the animal [3]; 2) the embryos of egg-laying fish can be readily cultured and staged throughout embryonic development, and the embryos are transparent, simplifying developmental studies. For these reasons I have undertaken the isolation and characterization of homeobox-containing genes from the teleost fish *Xiphophorus maculatus*.

When a *Xiphophorus* genomic Southern blot is probed with homeobox-containing fragments from the *Drosophila* homeotic genes *Antennapedia* (*Antp*) and *lushi tarazu* (*ftz*), several common hybridizing bands are detected, indicating the presence of a number of homeobox-related sequences within the fish genome (Fig. 1a). The *Xiphophorus maculatus* library (see Materials and Methods) was screened under identical conditions as for the genomic Southern with both *Drosophila* probes. Plaques hybridizing on both filters with both probes were picked and clonally purified.

As an independent assay for the relationship of these clones to *Drosophila* homeobox-containing genes, the hybridizing fragment from fish clone 2-1 was gel-purified and used as a probe onto a filter containing Eco RI digested *Drosophila melanogaster* DNA. As is seen in Figure 1b, the characteristic 'homeobox-ladder' (Fig. 1b, box) of hybridizing sequences, as reported by several groups [5,12], is detected. This strongly supports the proposal that at least clone 2-1 does contain a homeobox sequence. Clones 1-3, 2-1, 2-2 and 2-3 have been subjected to preliminary restriction enzyme mapping in conjunction with hybridization analyses. Figure 1c presents a representative result with one of these clones (1-3) and illustrates the strength and the specificity of the hybridization signal detected with a *Drosophila Antp* probe.

Figure 2a presents the results of *Antp* hybridizations for the four clones following digestion with various enzymes used in the restriction mapping. Clones 1-3 and 2-2 share one hybridizing band in common in the Pvu II, Ava I and Bgl II digests, suggesting that these two clones overlap and contain a homeobox sequence in common. This is shown more clearly in Figure 2b, where the hybridizing fragments from these two clones have been purified and used as hybridization probes onto fish genomic Southern blots. The 2-2 2.8kb, 1-3 2.3kb and 1-3

1.85kb probes all give rise to very similar hybridization signals, indicating that these fragments all derive from the same homeobox sequence. The clone 2-2 1.8kb fragment, however, gives rise to a completely different pattern of hybridizing bands, indicating that this fragment derives from an independent homeobox sequence. The actual relationship between clones 1-3 and 2-2 is shown in Figure 3, where it is seen that clone 2-2 does indeed contain two separate homeobox sequences, one of which is found in the partially overlapping clone 1-3. This result indicates that in *Xiphophorus*, analogous to the situation concerning homeobox-containing genes in *Drosophila*, *Xenopus*, mouse and man [2,4], homeobox-containing genes are clustered in the genome.

The homeobox region of clone 2-3 has been sequenced, and this is presented in Figure 4, along with a table of sequence comparisons to a number of other homeobox genes. The homeobox genes of *Drosophila* fall into several distinct classes, including the *Antennapedia* class (eg. *Antp* and *ftz*) and the *engrailed* class. Figure 4 shows that the homeobox in clone 2-3 has clearly strongest sequence identity to the *Antennapedia* class of homeoboxes. Because the hybridization conditions selected for screening the library with the *Antp* and *ftz* probes were probably not sufficient to detect cross-hybridization with *engrailed*-class sequences (W. Gehring, pers. comm.) the library has been rescreened with a *Drosophila engrailed* probe and positive clones have been picked and purified. Restriction mapping and sequencing of these clones is presently underway. In addition, Northern analyses are being carried out to ascertain when during development and in which (if any) adult tissues these genes are being expressed. In the near future in situ hybridizations will more precisely define the temporal and spatial pattern of mRNA expression for these genes. As the teleost fish allows ready access to all stages of vertebrate embryonic development, it is hoped that these studies will shed light onto what role(s) these genes play in vertebrate development.

REFERENCES- Appendix 5

1. Akam, M. "The molecular basis for metameric pattern in the *Drosophila* embryo" *Development* 101: 1-22 (1987)
2. Colberg-Poley, A. M., Voss, S. D., Chowdhury, K., Stewart, C. L., Wagner, E. F. and Gruss, P. "Clustered homeo boxes are differentially expressed during murine development" *Cell* 43: 39-45 (1985)
3. Fjose, A., Molven, A. and Eiken, H. G. "Molecular cloning and characterization of homeobox-containing genes from Atlantic salmon" *Gene* 62: 141-152 (1988)
4. Gehring, W. J. and Hiromi, Y. "Homeotic genes and the homeobox" *Ann. Rev. Genet.* 20: 147-173 (1986)
5. Hart, C. P., Awgulewitsch, A., Fainsod, A., McGinnis, W. and Ruddle, F. H. "Homeo box gene complex on mouse chromosome 11: molecular cloning, expression in embryogenesis, and homology to a human homeo box locus" *Cell* 43: 9-18 (1985)
6. Hogan, B., Holland, P. and Schofield, P. "How is the mouse segmented?" *TIG* 1:67-74 (1985)
7. Joyner, A. L., Kornberg, T., Coleman, K. G., Cox, D. R. and Martin, G. R. "Expression during embryogenesis of a mouse gene with sequence homology to the *Drosophila* engrailed gene" *Cell* 43: 29-37 (1985)
8. Joyner, A. L. and Martin, G. R. "En-1 and En-2, two mouse genes with sequence homology to the *Drosophila* engrailed gene: expression during embryogenesis" *Genes & Devel.* 1:29-38 (1987)
9. Keynes, R. J. and Stern, C. D. "Segmentation and neural development in vertebrates" *TINS*, May, 1985. pp 220-223
10. Lewis, E. B. "A gene complex controlling segmentation in *Drosophila*" *Nature* 276: 565-570 (1978)
11. McGinnis, W., Garger, R., Wirz, J., Kuroiwa, A. and Gehring, W. "A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans" *Cell* 37: 403-408 (1984)
12. McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. and Gehring, W. J. "A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complexes" *Nature* 308: 428-433 (1984)
13. Nüsslein-Volhard, C. In: "Determinants of Spatial Organization" S. Subtelny and I. R. Königsberg (eds.) Academic Press, N. Y. 1979. pp 185-211
14. Nüsslein-Volhard, C. and Wieschaus, E. "Mutations affecting segment number and polarity in *Drosophila melanogaster*" *Nature* 287: 795-810 (1980)
15. Ruddle, F. H., Hart, C. P. and McGinnis, W. "Structural and functional aspects of the mammalian homeo-box sequences" *Trends Genet.* 1: 48-51 (1985)
16. Schofield, P. N. "Patterns, puzzles and paradigms: the riddle of the homeobox" *TINS* 10:3-6 (1987)
17. Scott, M. P. and Weiner, A. J. "Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax and fushi tarazu loci of *Drosophila*" *Proc. Natl. Acad. Sci. U.S.A* 81: 4115-4119 (1984)
18. Stern, C. D. and Keynes, R. J. "Spatial patterns of homeobox expression in the developing mammalian CNS" *TINS* 11: 190-192 (1988)
19. Wakimoto, B. T., Turner, F. R. and Kaufman, T. C. "Defects in embryogenesis in mutants associated with the Antennapedia gene complex of *Drosophila melanogaster*" *Dev. Biol* 102: 147-172 (1984)

APPENDIX 5

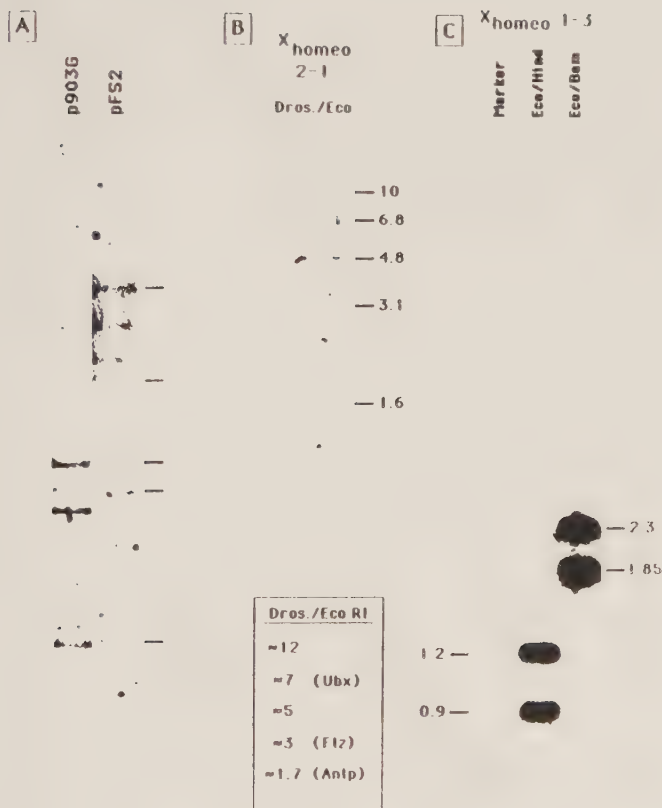


FIGURE 1

A) *Xiphophorus maculatus* genomic Southern probed with: p903G - a homeobox-containing fragment purified from a plasmid containing part of the *Drosophila Antennapedia* gene, and, pFS2 - a homeobox-containing fragment purified from a plasmid containing part of the *Drosophila tushi tarazu* gene. Hybridizing bands detected by both probes are marked. Plasmids p903G and pFS2 were kindly provided by W. Gehring, Basel.

B) Eco RI-digested *Drosophila melanogaster* DNA hybridized under standard stringency conditions with a homeobox-containing fragment from *Xhomeobox* clone 2-1. The hybridizing bands are marked with their size (in kb). In the box below are marked the Eco RI hybridizing fragments detected in *D. melanogaster* Southern with a *Drosophila* homeobox probe under reduced stringency conditions.

C) Hybridization of *Xhomeobox* clone 1-3 with the *Drosophila Antennapedia* probe (p903G).

APPENDIX 5

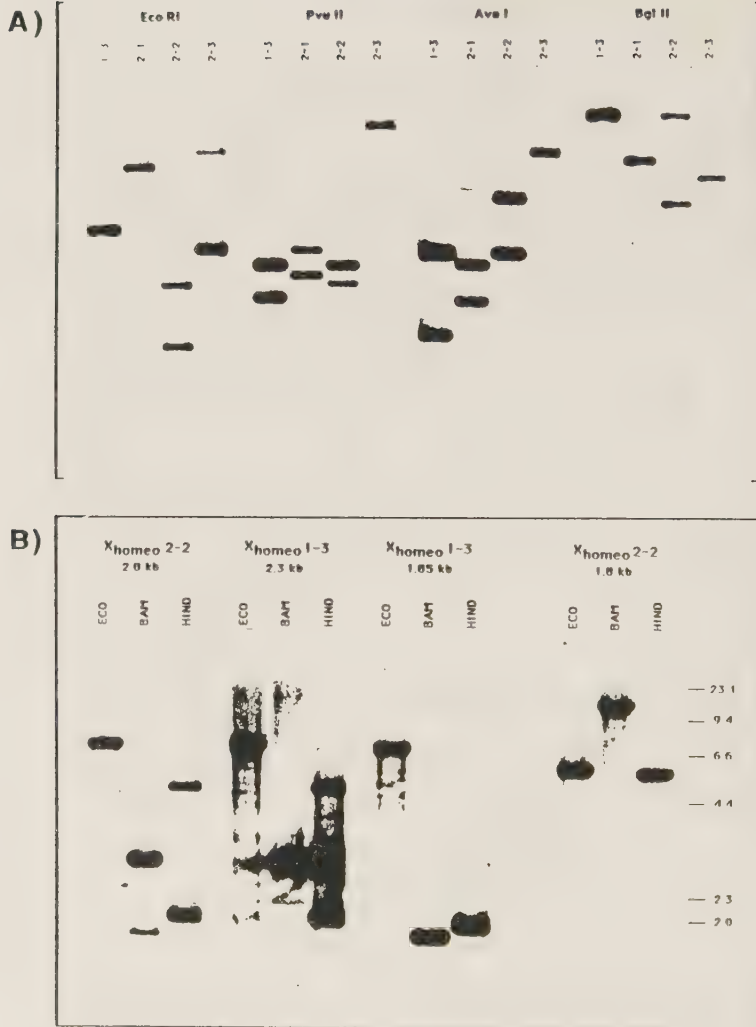


FIGURE 2

A) Digestion of four of the *Xhomeobox* clones with various restriction enzymes and hybridization with a *Drosophila Antennapedia* homeobox-specific probe.

B) Genomic Southern blots of *Xiphophorus maculatus* DNA probed with purified fragments from *Xhomeobox* clones 2-2 and 1-3. The 2.8 kb Eco/Bam and 2.3 kb Bam/Bam fragments from clone 1-3 contain the same homeobox (see also Figure 3).

APPENDIX 5

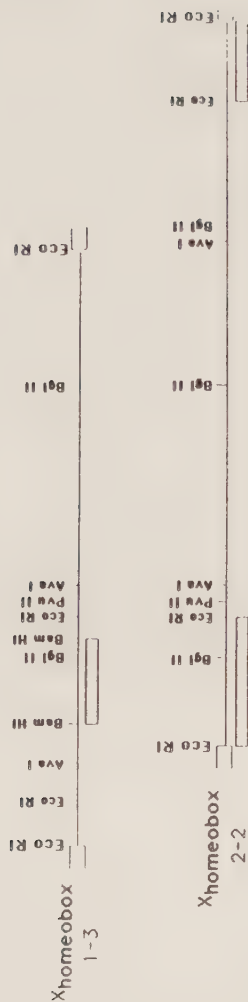


FIGURE 3

Partial restriction maps of *Xhomeobox* clones 1-3 and 2-2. These two clones overlap and demonstrate that, to some extent at least, homeobox-containing genes are clustered in the *Xiphophorus* genome.

APPENDIX 5

SALMON	G R G R G Q T I T R Y Q T L E L E X Z Y H F N R Y L T R R R R I E M A H A L C L S E R Q I K I V F Q N R R R H I V K I D N
SALMON	R E P G S Q T I T R Y Q T L E L E X Z Y H F N R Y L T R R R R V E I A H V L C L T E R Q I K I V F Q N R R R H I V K I D H
SALMON	P G I R I L R T A I T N T I Q L L E L E X Z Y H F N R Y L T R R R R V E I A A L D L T E R Q I K I V F Q N R R R H I V K I D R
ZEBRAFISH	G I R I S I R T S I T R Y Q T L E L E X Z Y H F N R Y L T R R R R I E I A N N C L M E R Q I K I V F Q N R R R H I V K I D S
XIPHOPHORUS	R X R I G I R I Q T T I S R R H I Q T L E L E X Z Y H F N R Y L T R R R R V E I A H A L S L S E R Q I K I V F Q N R R R H I V K I D H
DROS. ANT P	R G R I G I R I Q T T I T R Y Q T L E L E X Z Y H F N R Y L T R R R R I E I A H A L C L T E R Q I K I V F Q N R R R H I V K I D N
MOUSE HOX-2.1	R G R I A I R I Q A I T R Y Q T L E L E X Z Y H F N R Y L T R R R R I E I A H A L C L S E R Q I K I V F Q N R R R H I V K I D N
DROS. Ubx	R R R I G I R I Q T T I T R Y Q T L E L E X Z Y H F N R Y L T R R R R I E M A H A L C L T E R Q I K I V F Q N R R R H I V K I D I
DROS. ENGRAILED	E K R I P I R T A P S S E Q L A R L E R E F F N E N R Y L T F Z R R R Q L S S E L G L N E A I Q I K I V F Q N R R R H I V K I D S T

AMINO ACID IDENTITY

X homeobox 2-3 versus:

Dros. Antp	86.7
Dros. Ubx	80.0
Dros. eng.	53.3
Mouse Hox-2.1	85.0
Zebrafish (ZF 2S)	80.0
Salmon (ps6)	63.3
Salmon (ps12-A)	90.0
Salmon (ps12-B)	81.7

FIGURE 4 Comparison of homeobox sequence of Xhomeobox 2-3 to several published sequences (Incl. fish)

IX LEBENS LAUF

Name:	Scott McGregor Robertson
Geburtsdatum:	20.10.1956
Geburtsort:	Auckland, New Zealand
Eltern:	Murray Spencer Robertson, M.D. Freda Mae Robertson, geb. Padgett
1962-1968:	Waihopi Elementary School, Invercargill, New Zealand
1968-1969:	McKittrick Junior High School, North Battleford, Sask., Canada
1969-1970:	Dalewood Junior High School, St. Catharines, Ontario, Canada
1970-1972:	Grantham High School, St. Catharines, Ontario, Canada
1972-1975:	Governor Simcoe Secondary School, St. Catharines, Ontario, Canada
1975-1980:	McMaster University, Hamilton, Ontario, Canada
1980:	Batchelor of Science (Honours) Degree in Hauptfach Biologie
1980-1983:	University of Toronto, Toronto, Ontario, Canada
1983:	Master of Science Degree in Hauptfach Biologie
1983-1985:	Justus-Liebig-Universität, Giessen, FRG
1985-1988:	Ludwig-Maximilians-Universität, München, FRG
03/1985- 08/1988	Promotion unter Anleitung von Dr. M. Scharf im Laboratorium für Molekular Embryologie (Genzentrum) der Max-Planck-Institut für Biochemie, Martinsried, FRG

Zu diesem Buch:

Die Isolierung verschiedener Vertreter der *src*-Genfamilie des Knochenfisches *Xiphophorus maculatus* erlaubte Untersuchungen

- über die Verwandtschaftsbeziehungen der Sequenz und der Gen-Topologie dieser Gene eines niederen Vertebraten zu den homologen Genen von Invertebraten und höheren Vertebraten;
- über die Beziehung des *c-src* Gens von *Xiphophorus* zum Gen-Locus *mdl*, der für die Melanombildung in Hybriden von Platyfisch und Schwertträger verantwortlich ist und
- über verschiedene Aspekte der Evolution der *src*-Genfamilie, einschließlich der Berechnung des Zeitpunkts der *src/yes* Gen-duplikation.

Zum Autor:

Scott Robertson studierte Biologie in Canada an den Universitäten Hamilton und Toronto sowie an der Justus-Liebig-Universität Gießen. Die Forschungsarbeiten zur vorliegenden Veröffentlichung wurden am Max-Planck-Institut für Biochemie in Martinsried bei München - unter der Leitung von Herrn Dr. M. Schartl im Laboratorium für Molekulare Embryologie (Genzentrum) - durchgeführt und von der Fakultät für Biologie der Ludwig-Maximilians-Universität München als Promotions-schrift angenommen.